DOCUMENT RESUME

ED 039 613 CG 005 636

AUTHOR	Marascuilo, Leonard A.
TITLE	The Effects of Heterogeneous and Homogeneous Groupings on Student Attitudes and Student Performance in Eighth and Ninth Grade Social Studies Classes.
INSTITUTION	California Univ., Berkeley.
SPONS AGENCY	Berkeley Unified School District, Calif.; Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-8-I-046
PUB DATE	1 Apr 70
GRANT	OEG-9-8-081046-0113-(010)
NOTE	216p.
EDRS PRICE	EDRS Price MF-$1.00 HC-$10.90
DESCRIPTORS	Ability Grouping, *Academic Performance, Attitudes, Class Organization, *Grouping (Instructional Purposes), Heterogeneous Grouping, Homogeneous Grouping, *Junior High School Students, *Racial Segregation, *Student Attitudes

ABSTRACT

This study attempts to determine whether or not ability grouping is desirable for effective instruction. Parents volunteered their children for assignment to a heterogeneous social studies class. Students were divided into a Volunteer Group, a Non-volunteer Group and a Non-response Group. Students for the heterogeneous classes were randomly selected from previous tracks and were distributed among four experimental classes, in a proportion representing the proportion of each group in the eighth grade population. The balance of the Volunteer Group were distributed into regular track classes. Other groups were similarly placed. Results of the factor analyses indicate that student attitudes are not affected by class type. Analyses of variance indicated that students from lower tracks achieved at a higher level in a heterogeneous grouping than in a homogeneous one. Teacher activities were similar with both groups, with student interaction apparently accounting for achievement differences. Sample student questionnaires and tests, as well as parent letters are included. (Author/CJ)

THE EFFECTS OF HETEROGENEOUS AND HOMOGENEOUS GROUPINGS ON STUDENT ATTITUDES AND STUDENT PERFORMANCE IN EIGHTH AND NINTH GRADE SOCIAL STUDIES CLASSES

Author of the Report:

Dr. Leonard A. Marascuilo
Associate Professor of Education
University of California, Berkeley

THE EFFECTS OF HETEROGENEOUS AND HOMOGENEOUS GROUPINGS ON STUDENT ATTITUDES AND STUDENT PERFORMANCE IN EIGHTH AND NINTH GRADE SOCIAL STUDIES CLASSES

Sponsored by Berkeley Unified School District
Under
U.S.O.E Grant No. OEG-9-8-081046-0113-(010)

Project Initiator:
Dr. Joseph N. Rodeheaver
Director of Secondary Education
Berkeley Unified School District

Project Director:
Mr. Edward A. Lewis
Counselor at West Campus
Berkeley Unified School District

Author of the Report:

Dr. Leonard A. Marascuilo
Associate Professor of Education
University of California, Berkeley

1 April 1970
Berkeley, California

Table of Contents

ACKNOWLEDGEMENTS

The evaluator would like to take this opportunity to thank some of the people whose help and assistance made this report possible. A study of the size and scope of the one reported in these pages requires the great interest and hard work of a large number of people. The success of this study is the result of this special kind of effort.

Special thanks are given to Dr. Joseph N. Rodeheaver of the Berkeley Unified School District who kept the project in motion from his offices; to Mr. Stanley Wright who prepared the proposal which was eventually accepted by the U. S. Office of Education, who served as Project Director for the first year of the study, and who is responsible for the history of testing and ability grouping which appears at the beginning of this report; to Mr. Edward A. Lewis who served as Project Director during the second year of the study; to Dr. Eugene McCreary, supervisor of teacher education in social studies at the University of California who served as a consultant to the teachers for the second year of the study; and to Miss Virginia Thickens, Curriculum Associate for History at West Campus, who served as coordinator for the social studies teachers.

Special thanks are also given to the following teachers who taught heterogeneous classes during the study and whose comments form a portion of this report: Mrs. Minerva Curtis, Mrs. Lorna Hall, Mrs. Fanny Smith, Mrs. Florence Jones, Mrs. Ellen Sipe, Mrs. Lana Hoch, Miss Susan Lom, Mr. Jack Millinger, and Mr. Robert Scrofani.

Thanks, too, to Miss Jean Hill, Miss Rachel Blau, and

Mr. Chris Lim, who were the secretaries and clerk assistants at West Campus during the project; to Mr. John Reed who took complete charge of the data processing for the two years of the study; to Miss Monica Arndt, who did all the key punching; and to Mr. Thomas W. Little, who typed the drafts and the final copy of this report.

THE EFFECTS OF HETEROGENEOUS AND HOMOGENEOUS GROUPINGS ON STUDENT ATTITUDES AND STUDENT PERFORMANCES IN EIGHTH AND NINTH GRADE SOCIAL STUDIES CLASSES

Author of the Report:

Leonard A. Marascuilo
Associate Professor of Education
University of California at Berkeley

Introduction

Urban schools face the problem of giving equal opportunity to children with a variety of backgrounds and abilities. Due to the increase in pupil population in large school districts educators have been forced to concentrate on supplying new school plants and have been unable to meet the challenge of finding and developing sound ways of educating children. On June 20, 1967, U. S. Circuit Court of Appeals Judge James Skelly Wright ordered that the track system in Washington, D. C.'s public schools be abolished immediately. In so doing he struck at an area in education which has been open to constant debate but about which no widely accepted conclusions have been reached.

Julius W. Hobson, whose suit was decided on June 20, 1967, says that school officials have deliberately deprived Negro and lower class students of adequate textbooks and facilities such as libraries. The decision did not accept this contention, but this should give educators no cause for complacency; as Susan Filson, Washington Post Staff Writer, stated: "School officials can and should be faulted, but for another reason: that in the face of those inequalities they have sometimes shown criminally little concern. It is one thing, for example, when crowded residential conditions shut Negro children, and them alone, out of kindergarten in the nearby schools. It is another matter entirely when, confronted with this awful situation, school officials stand idly by, circulating promises of more adequate school buildings years hence, but acquiescing in the present injustice by their passivity."

In many school districts throughout the nation we have hundreds of examples of educators in the quandary of what to do about tracking. In practice, if not in concept, states the court decision, the track system discriminates against the disadvantaged child, particularly the Negro. Designed in 1955 as a means of protecting the school system against the ill effects of integrating with white children the Negro victims of de jure separate but unequal education, it has survived to stigmatize the disadvantaged child of whatever race relegated to its lower tracks--from which tracks the possibility of switching upward, because of the absence of compensatory education is remote.

It is common knowledge that many school systems in our states use a form of ability grouping in which students are divided into separate, self-contained curricula or tracks ranging from Basic for the slow student to Honors for the gifted. As might be expected, all fifty states use some form of grouping. This is what Wright ordered abolished in the Washington, D. C. schools.

In most communities schools are racially and socially homogeneous and many so-called integrated schools are actually racially segregated within the school. Educators have known for some time that these conditions damage the minds and spirit of all children who attend them--the Negro, the white, the poor, and the affluent--and block the attainment of the broader goals of democratic education, whether the segregation occurs by law or by fact. Knowing all this it seems surprising that educators

have not taken steps to abolish ability grouping and bring about real integration. However, a strong opinion prevails among teachers and administrators that ability grouping is necessary for effective instruction. Whether it is actually necessary or desirable for effective instruction is what this study will seek to determine. A survey of the growth of testing will indicate how the opinion of teachers and administrators slowly formed over the past sixty-seven years and how it influenced the trend toward ability grouping.

History of Student Testing and Its Import on Ability Grouping

The first 67 years of the twentieth century may be conveniently divided into four parts. We may designate the period from 1900 to 1915 as the pioneering phase. This was the period of exploration and initial development of methods. It saw the emergence of the first Binet intelligence scales and the American revisions. Standardized achievement tests in different subjects began to appear, exemplified by Stone's arithmetic tests, Buckingham's spelling tests, and Trabue's language tests. Otis and others were initiating work on group tests of intelligence.

The next 15 years, 1915 to 1930, can be called the boom period in test development. The pioneers had shown the way, and in the hands of enthusiastic followers tests multiplied. Standardized tests and achievement batteries made their appearance. Tests of intelligence and achievement were administered widely and somewhat indiscriminately. Test results were often accepted unhesitatingly and uncritically and served as the basis

for a variety of unjustified judgements and actions with respect to individuals.

A critical attack by some educators had the healthy effect of forcing the test enthusiasts themselves to become more critical of their assumptions and procedures and to broaden their approach to the whole problem of psychological and educational appraisal. Meanwhile in the field, educators were using the tests in their own way to further education. From about 1930 to 1945 may be considered a period of critical appraisal, of taking stock, of broadening techniques and delimiting interpretations. It was a period in which the center of attention shifted from measuring a limited range of academic skills to evaluating achievement of the whole range of educational objectives. It was a period in which the holistic, global projective methods of personality appraisal came to the fore.

From 1945 to the present we have had a series of test batteries and testing programs. We have experienced a second boom period--not so much in test development and construction as in test administration and use. The mid-twentieth century is a period in which standardized testing is a widely experienced and widely accepted phenomenon of our American culture. Now we have to evaluate our motives when we test. There is nothing wrong with the tests per se, but our uses of them are leading us steadily into trouble.

At the same time that tests were being developed studies were conducted to attempt to evaluate the uses and particularly the abuses of the tests. One of the abuses of the tests has been the ability grouping of children.

One of the best historical reviews of ability grouping was carried out by Wyndham[4]. It may be regarded as a definitive review of the position in respect to class grouping in the United States until 1933. The review falls into two sections. The first examines the stages, between 1870 and 1920, in which school practice broke away from an acceptance of a uniform curriculum organized grade by grade, with attainment as the sole criterion of promotion. The first stage was one in which a few innovators, without questioning the sanctity of the curriculum or attainment as the criterion of pupil progress, sought a measure of administrative elasticity in a greater frequency of promotion of individuals or groups of pupils. This meant the relative acceleration of some pupils through a standard curriculum. By the turn of the century there were a few attempts to provide for apparent differences in ability by a modification of the curriculum for groups or for individual pupils. The next stage was characterized by the growth of medical inspection of school children, resulting, among other things, in a dominant interest in pupils backward in attainment in terms of age. "Retardation" became a familiar term. Ayres' Laggards in Our Schools provided the material for debate.

In the next stage the interest of psychologists in mentally defective children and the development of intelligence tests not only revealed further deficiencies in school practices, but also armed critics with weapons of a new type. No widespread attack upon the general problem of class grouping had been made.

The second section of the review covers the period

1920 to 1932. By 1920 many long-cherished assumptions as to curriculum and method were being challenged. It is significant that many terms now long familiar date from the years 1919-1920: Dalton Plan, progressive education, "the child-centered school," individualization of instruction, the Winnetka Plan. Meanwhile use was being made, to an increasing extent, of tests of intelligence and of attainments for purposes of class-grouping. The review examines the extent to which "homogeneous grouping" began to appear as an administrative device during this period and the nature of the criteria employed. An account is given of the growing debate on "ability grouping." (1) An examination of school practices indicated that ability grouping was never adopted during the period under review to the extent assumed by many of the debaters on both sides. A critical examination of the research studies carried out between 1920 and 1932 yields disappointing results if one is looking for final answers to any of the questions involved. The reason lies first in the fact that certain assumptions must be made as educational purposes. For example, one must determine whether a homogeneous group is to be regarded as an end in itself or whether it is possible to think of it as a permanent state of affairs. Secondly, the human situation of teacher and class involves variables which elude statistical control or analysis, and some of these variables may well prove of the greatest significance from the point of view of educational aims and values. Thirdly, even when they were limited to aspects of the problem which might well be considered amenable to controlled experimentation, a number of the studies were vitiated by faulty experimental

design or lack of adequate statistical control and analysis. The study provides a detailed analysis of the most significant of the experimental studies of this period, and while the results provide no unequivocal answers, the questions raised then still remain to be answered. In the United States, research findings are still inconclusive. (2) In the U. S., this is doubtless true because the center of educational debate has moved to other topics. In England the debate on "streaming" has been conducted in many centers more in terms of social issues than of experimental results.

Yet, since school authorities must provide for a large school population with a teaching force which is not limitless, the problem of the most desirable means of grouping children in school remains. Assuming that a curriculum is constructed which provides for differences in abilities and needs, it has yet to be determined to what extent instruction, with the aid of modern devices, should be on an individual basis and for what purposes and on the basis of what criteria those individual pupils should be brought together for corporate experience.

Ruth B. Ekstrom[2] takes over from Wyndham and states the controversy in her review of the literature from 1930 to 1959. James B. Conant has recommended ability grouping in required high-school subjects and in subjects elected by students with a wide range of ability[3]. According to Otto[4]; the evidence slightly favors the use of ability grouping, particularly where standards, materials, and methods are adapted to the group. The greatest relative effectiveness of homogeneous grouping is

indicated for dull children, the next greatest for average children, and the least benefit and most frequent harm for bright children. Other reviewers have felt that the experimental evidence used to compare homogeneous and heterogeneous grouping has been inadequate, but they have pointed out the importance of adapting teaching methods and curriculum to the pupils[5,6,7,8,9]. Ekstrom concluded that the major deficiency in most of the experiments that compared homogeneous and heterogeneous grouping is failure to provide for differentiation of course content and method according to level of ability. There is a great variety of experimental design and no consistent pattern of results. Many experiments failed to control the type of teaching and to provide differentiation of teaching according to ability levels. Poor experimental design, such as the use of available data only and the use of matched pairs of subjects on unwarranted assumptions of similarity, made many studies less effective. In experiments that specifically provided for differentiation of teaching methods and materials for groups at each ability level, and made an effort to push bright homogeneous classes, Ekstrom states that results tended to favor the homogeneous classes.

From 1958 on, some of the most significant studies in the effects of grouping have taken place. Luttrell[10] attempted to compare the academic achievement and personality development of 27 gifted sixth graders in a special class with that of a comparable group in eight regular classes. He found that the academic achievement for the special gifted class was significantly greater than that for gifted in the regular classes. In

contrast to this Schaefer[11] said that he believed it to be a mistake to segregate the so-called superior students. Koontz[12] attempted to assess the effects of grouping on achievement when curriculum and instruction are modified to suit the varying levels of pupil ability. He found that homogeneous grouping failed to realize its theoretical possibilities.

Drews[13] tried to determine the effect of homogeneous and heterogeneous grouping on children at three ability levels in ninth-grade English classes. She found that homogeneous classrooms appear to have little advantage of heterogeneous in meeting the needs of students as judged by teacher, peer, and self ratings. In view of this study there seems to be a need for further study of criteria of selection for grouping and their efficiency in the reduction of individual differences in intelligence and achievement at the classroom level.

Atkinson and O'Connor[14] initiated two studies to explore some of the potential effects of ability grouping on motivation and the effects of homogeneous ability grouping in the sixth grade. The outcomes of this study of sixth-grade pupils indicate that placement in homogeneous groups does not lead to a general enhancement or decline in interest in learning. They found that motivational dispositions interact with treatment so that the effect for some students is advantageous and for others disadvantageous. Students high in resultant achievement motivation show higher growth in achievement and greater interest in schoolwork when in homogeneous groups than in heterogeneous groups. Students low in resultant achievement motivation (the more anxious students) show a decline in interest

when placed in homogeneous ability groups but no marked difference in achievement. Within homogeneous classes achievement motivation is positively related to better growth and higher interest. Within heterogeneous classes there is no relationship between motivation and interest or growth in performance between fifth and sixth grades. Differences associated with motivation are obtained both for students of high ability placed in special advanced sections and for students of lower levels of ability placed in regular sections. If these results are confirmed in subsequent investigations, achievement motivation should become a pertinent factor in determining which children should be assigned to homogeneous ability groups if maximization of interest in schoolwork and learning is desired. Borg[15] attempted to study differences in the effects upon elementary, junior high, and senior high school pupils of an ability grouping system that differentiates the curriculum principally by adjusting the rate of presentation of curricular materials and a random grouping system that differentiates the curriculum principally through the use of enrichment. He found that when all differences between ability grouped and random samples are considered, the most tenable conclusion to draw from the data is that neither ability grouping nor random grouping has a consistent general effect upon achievement at any of the grade levels tested. He concluded that superior pupils generally develop superior study habits in both grouping treatments and counters the frequently heard criticism that superior pupils do not learn to study in the public schools.

One of the best statistical studies on the effects of

ability grouping was carried out by Goldberg, Passow, and Justman[16]. They tried to explore the differences in achievement and learning patterns, social and personal relations, motivational structure, and attitudes toward self and toward school of intermediate grade children when grouped in classes with various ranges of intellectual ability. They used the hypothesis that neither the range of abilities in any given classroom nor the presence or absence of gifted pupils will affect the attainments of elementary school pupils.

The findings were that for all the variables studied, the effects of administrative grouping were, at best, minimal. Certainly, differences in achievement growth over the two grade span did not support the common wisdom that narrowing the ability range or separating out the extreme groups from the intermediate groups enables teachers to be more effective in raising the pupils' achievement level. The evidence from this study suggests that such is not necessarily the case. Not all teachers took advantage of the narrower range to do the very things which such grouping should have made possible.

Although the achievement differences among ranges were small, overall observed increments tended to favor the broad range situations. The findings cast serious doubt on the effects of ability grouping per se in raising the academic attainment of pupils.

Finally, they state that in the absence of specific plans for changing the content and methods of teaching so as deliberately to provide the most needed and challenging learning situation for each group of pupils, ability grouping does not

seem to make any appreciable difference.

Dockrell[17] studied the cognitive outcomes of homogeneous versus heterogeneous grouping at the end of a two year study. Before this study most junior high school pupils were assigned to class groups within a grade without regard to achievement level or ability though one or two schools were beginning to experiment with some form of achievement grouping.

He found that in comparing different achievement levels most significant differences occurred for the lowest level. The lowest achieveing pupils did better in heterogeneous classes, and "streaming" was irrelevant for middle and high achievers. This was not a laboratory experiment under ideal conditions but a field study under conditions as they exist in the Edmonton schools. Teacher variables were controlled, differentiated curricula and teaching materials provided and a full range of socio-economic groups were included. It seems unlikely, therefore, that under present circumstances these conclusions could be materially modified.

The function of this survey has been to examine some of the research and informed judgment in regard to a study on homogeneous and heterogeneous grouping in eighth and ninth grade social studies. The studies point out that:

1. It is not easy to determine what children learn, how much, how well, and toward what ends.

2. Attempts to find out the extent to which certain grouping situations contribute to a child's progress or lack of it pose a number of new questions.

3. The significance of diverse kinds of learning

gains that result from different ways of grouping should be a subject of continuous study.

4. Factors other than the particular grouping methods used may account for differences that show up in achievement gains between children grouped according to ability and those grouped heterogeneously.

5. There is great difficulty deciding if grouping children is educationally sound.

The research studies indicate a tentative conclusion: On the average, achievement gains made by pupils in classrooms representing more than a normal spread of differences among children were higher than average gains made by pupils in ability-grouped classes.

There is a special area which most of the studies have not even considered. In a recent (1966) racial and ethnic survey of California public schools it was brought dramatically to the attention of the community that one in four of the children enrolled in California schools is a member of a minority group. Of 1,276 schools in the state's eight largest school districts--all urban or suburban and each with an enrollment of more than 50,000--419 were found to be mixed, 440 to have a high concentration of minorities, and 417 to have a high concentration of the Anglo majority group.

Dorman L. Commons, president of the state board, commented that California has not been "walking fast enough to keep up with itself" in its efforts to integrate schools and that the survey results comprise a "frightening report." Unless studies are undertaken many vital decisions are going to

to be made on educated guesses--and that is exactly what has landed us in our present state of indecision. Coleman[18] says that of all the characteristics of schools which distinguish the education being provided the average white and Negro students, it is the environment provided by the fellow students where the differences are most dramatic.

Description of Grouping in Berkeley Schools

At the time of the study, the then current practice in the Berkeley Unified School District junior and senior high schools was to form four broad groups which were homogeneous with respect to a rather loosely defined combination of academic ability and achievement in the given subject. At the beginning of the seventh grade the students were placed in groups by the counselor on the basis of the sixth grade teacher's recommendation and with the help of standardized tests (the Stanford Achievement Test and the Lorge-Thorndike Intelligence Test). After the first grading period there was considerable reassigning of the students to groups. In fact, all through a student's school career he may have had his group assignment changed if there was evidence that he was misplaced.

In the junior high schools, English, history, science and mathematics were grouped into four groups. (In addition, classes in art, music, and foreign languages were divided, although along slightly different lines.) Here is a qualitative summary of the seventh and eighth grade groups in social studies in the 1967-1968 school year. The top group was college preparatory and had about 35 percent of all of the students. Its racial composition was overwhelmingly Caucasian and Oriental.

Group Two spanned the middle of the range of abilities and was also college preparatory. It contained about 40 percent of the students. Its racial composition was about that of the whole student body: i.e., about 50 percent Caucasian, 40 percent Negro, and 10 percent Oriental. Group Three was not college preparatory. It consisted of students working considerably behind grade level and contained about 20 percent of the students. Its composition was 75 percent Negro. Group Four contained pupils deemed incapable of participating in Group Three classes. It had about five percent of the school population and was overwhelmingly Negro. Special classes for the educationally handicapped and educationally mentally retarded were offered to those students who qualified.

Some Unintended Results of the Grouping System in the Berkeley Schools

1. De facto racial segregation. The most obvious unintended result of the grouping system in the Berkeley Unified School District (BUSD) is that it produces de facto segregated classrooms within the integrated secondary schools. Racial stereotyping, social circumstances and differences in educational background result in the top group being populated almost exclusively by Caucasians and Orientals and the bottom groups almost exclusively by Negroes.

2. Intellectual Isolation. The main idea behind grouping in the BUSD schools is the separation of the students according to intellectual or academic ability, so that the teacher can teach the separated groups more efficiently. Unfortunately, in addition to the possible benefits of this separa-

tion, there are some unintended side effects which result from such a structure. According to the recent (1966) Federal study described in the Coleman Report, "Equality of Educational Opportunity," the factors found to correlate most strongly with high scholastic achievement are the quality of educational background and the level of educational aspirations of one's classmates. It may follow from this empirical result that a large part of the very plausible theory lying at the base of the grouping system may be wrong. According to this theory, the best way to teach the slow students is to put them all together and, to put it simply, talk slowly to them. According to the Coleman Report and according to experiences in BUSD and elsewhere, it doesn't work that way in practice. A student's classmates play a very important role in determining his intellectual progress.

That this should be so makes a great deal of sense, since one of the causes of low achievement is lack of motivation and aspiration. It is unreasonable to expect a low-achieving student to pick up more favorable attitudes if his classroom contacts are limited to other students with low motivation and aspirations.

3. Cultural and economic isolation. It is well established that a child's performance in school (and on tests) is correlated to his cultural background and to the income level of his family. Generally speaking, middle and upper income families have parents whose educational level is higher than the parents of lower income families. The educational background of the parents affects the achievement and atti-

tudes of the children. The attitude with which a child approaches an IQ or aptitude test varies a great deal with his cultural background. Furthermore, the tests themselves take for granted a certain cultural background. As a result of these factors, the students who do well on the tests and who do well in school--and are consequently grouped together--tend to be from similar backgrounds. The same is true of those who do poorly. Thus, the grouped classes tend to be homogeneous with respect to cultural and economic factors as well as intellectual factors.

There are many citizens who believe that contemporary society suffers greatly by virtue of its being separated into isolated economic and cultural sub-communities. Grouping tends to perpetuate this separation and isolation. Effectively, although unintentionally, it is as if the youngsters from a particular sub-community were fed into particular groups which later feed them back into the same sub-community. Instead of the school's tending to unite the diverse elements of society, it thus serves to solidify and perpetuate the separation and isolation.

It is felt that the diversity found in the BUSD community is itself an educational asset, and that separation and isolation of diverse groups in the community is as harmful to the white majority as it is to the Negro minority. America <u>is</u> in the midst of a profound social revolution. Its children will have a better understanding of their country, as well as of their community, the world, and themselves, if they learn to feel at home with the broad spectrum of individuals and back-

grounds within its own city boundaries. Many adults feel uncomfortable in the presence of individuals who do not share their skin color or cultural heritage. This uneasiness of the adults is to some extent a result of the fact that their contacts as children were limited to like-minded and like-looking people.

4. Lowering of aspiration. Another effect of the grouping system results from the fact that status, high or low, is attached to an individual on the basis of the groups he is placed in. The worst aspect of this is that a child in the lower group may suffer a crushing blow to his self-esteem by being labeled "low group." His aspirations also tend to fall because not much is expected of him. He then accomplishes less than he is capable of accomplishing.

5. Stereotyping of students. Parents expect teachers to be sensitive to the individual needs of their children, and to teach them in such a way that these individual needs are met and individual talents developed. It is more difficult for a teacher to accomplish this aim when each child is remanded to him as part of a group which is presumed to have many characteristics in common, characteristics which may or may not hold for a particular child. It is difficult for a teacher to retain his sensitivity to the individuality of a child if someone has already classified the child as part of a group. It is a temptation for the teacher to assume that the child's individual needs are being sufficiently met by the grouping practice itself, thereby relieving the teacher of further responsibility for adapting his teaching to the individual child.

In most communities, the possible transition to heterogeneous classes will involve a great deal of careful planning and preparation. The detailed manner in which the transition is accomplished is the responsibility of the administration and the staff. This study was undertaken to help school administration decide whether homogeneous classes are indeed necessary for quality education in America's public schools.

Reasons for this Study

The reasons for this study were put forth in the previous sections. In addition, other objectives were of interest. They are:

1. To find out if students learn more in a heterogeneous classroom setting than in a homogeneous classroom setting.

2. To help administrators and teachers structure a quality educational program attuned to the demands of best educational practices.

3. To help create an educational program responsive to the needs of the Berkeley community.

4. To show the community that the BUSD decision to integrate pupils is a responsible, well-thought-out, statistically valid, non-emotional progressive act.

Procedure Used to Define the Experimental Groups of the Study

At the end of the spring semester of 1967, a questionnaire was distributed to all parents and guardians of students assigned to the state required classes in eighth grade social studies. In the questionnaire, parents were informed of the intent of the school district to perform the study described

in this report. Each parent was invited to participate in the study by volunteering their children for assignment to a heterogeneous social studies class for the upcoming school year.

On the basis of the parent response the entire entering eighth grade was divided into a Volunteer Group, comprising students whose parents expressed a willingness for the participation of their children in the experimental heterogeneous classes, a Non-Volunteer Group, comprising students whose parents expressed opposition to participation in the experimental classes, and a Non-Response Group, comprising students whose parents made no reply to the question of whether they approved of their children's participation or opposed it.

From the Volunteer Group, the students for the heterogeneous classes were randomly selected from social studies Track One, Two, and Three. These students were randomly and evenly distributed among four experimental classes, in a proportion which represented the proportion of each group in the eighth grade school population as a whole; that is, for each experimental class, 10 students were from Track One, 11 students were from Track Two, and seven students were from Track Three. The balance of the Volunteer Group were evenly and randomly distributed into their regular tracked classes. These Volunteer Groups can be considered as control tracks designed to eliminate any bias inherent in the act of volunteering.

The Non-Volunteer Group was evenly and randomly distributed across the projected tracked classes and the Non-Response Group, also, were evenly and randomly distributed across the ordinary tracked classes. Efforts were made to maintain a pro-

portionate balance representative of the whole grade between boys and girls in each distribution, though it is obvious that this was not possible in groups where there were normally more boys than girls or where the opposite condition existed. In any case, the proportion of boys to girls in each class was approximately equivalent to the normal distribution of these groups.

The number of classes available for the first year's study was as follows:

Track One	6 classes of size 32
Track Two	7 classes of size 32
Track Three	3 classes of size 25
Experimental Heterogeneous	4 classes of size 28
Track Four	3 classes of size 17
Honors	1 class of size 20

As the breakdown and distribution suggests, each heterogeneous class was composed of approximately 10 Track One students, 11 Track Two students, and seven Track Three students. This means that each heterogeneous class had a fairly large group of high ability students, since it must be recalled that many of the Track Two students are college potential as are most of the Track One students. In the interpretation of results, this must be taken into consideration. The schedule of classes for the day was essentially as shown in Figure 1.

Statistical Procedures Used to Process Data

The object of the analysis is to compare and ascertain differences that may be manifested between and within classes on the criterion measures. Analyses of variance are

Figure 1. Teacher and Class Schedule by Periods for the First Year's Study.

Period	Teacher I	II	III	IV	V
1		Hom* Track 3	Hom Track 2	Track 4	
2				Hom Track 2	Hom Track 1
3	Het**	Het	Hom Track 3		
4		Hom Track 2	Hom Track 3	Hom Track 1	Track 4
5	Honors				Hom Track 1
6				Hom Track 2	Hom Track 2
7	Hom Track 1	Hom Track 2			
8	Hom Track 2	Hom Track 1	Het	Het	
9	Hom Track 3		Hom Track 1		Track 4

*Homogeneous classes consisting of Volunteer, Non-Volunteer, and Non-Response Group students.

**Heterogeneous classes.

used to determine if statistically significant differences exist at the end of the experimental terms and school years. Appropriate post hoc comparisons and confidence intervals using Scheffé's method of multiple comparisons are used to assess the value of significant findings.[19] The comparisons of interest are between the students of each group who are in the experimental heterogeneous classes and those of the same group who are in the ordinary homogeneous classes. Comparisons are also made to determine if students whose parents have indicated a willingness for them to participate in the experiment are different on the criterion measures from those whose parents have opposed participation or who have not responded. By distributing the classes evenly among the teachers, as far as this was possible, the effects of teacher differences are hopefully controlled in the analysis.

For the analysis of attitude data, factor analyses were performed to create scale values. These scale values were then used in analyses of variance designs to study the basic hypotheses of the study. Stated in positive form, these hypotheses are:

1. Among the various ability grouped children placed in the heterogeneous classes, Track One children will do as well as the Track One children in the homogeneous classes.

2. The Track Two children placed in the heterogeneous classes will do significantly better than those Track Two children who are placed in the homogeneous classes.

3. The Track Three children placed in the heterogeneous classes will do significantly better than those Track

Three children who are placed in the homogeneous classes.

Description of the Experimental Groups for the 1967-1968 Study

From among the students whose parents had indicated willingness for them to participate in the experimental heterogeneous classes for social studies, groups in the general ratio of 10 Track One students to 11 Track Two students to seven Track Three students were selected by use of a random numbers table to form experimental heterogeneous classes at Garfield Junior High School. The remainder of the students who had volunteered but could not be accommodated in the small number of heterogeneous classes were evenly and randomly distributed across the regularly tracked homogeneous classes. All of the students whose parents had either declined to have them participate or had failed to respond to the request for student participation were randomly distributed to form regularly tracked classes. Thus, for each normal track designation there resulted three kinds of students: Volunteers in the heterogeneous classes, non-selected volunteers assigned to homogeneous classes, non-volunteers assigned to homogeneous classes. Thus, the design for the comparisons summarized here was as shown in Figure 2.

As a starting point, it seemed reasonable to assume that students whose parents were concerned about their children's performance and/or about the possible effects of heterogeneous grouping upon school learning and social development would also be subject to differences in either home environment or home attitudes. Therefore, the most important comparisons for the evaluation were designated as the differences between the volunteers in the experimental classes and the non-selected volun-

Figure 2. Basic Design of the Study for the 1967-1968 School Year.

Track	Heterogeneous Volunteers	Non-selected Volunteers in Tracks	Non-volunteers
One			
Two			
Three			

teers who had been distributed across the regular homogeneous classes. It is believed that the comparisons between these two groups holds constant any differences due to the forces operating on parents and students which resulted in the event "volunteering for the study."

Initial Differences in School Achievement and Ability of the Experimental Groups

Results from the regular testing of the eighth grade students on the Sequential Tests of Educational Progress (STEP) and the School and College Ability Test (SCAT), administered in the fall of 1967 were used to determine if differences in ability existed between the three groups or kinds of students within each track. For the analysis, scores were standardized so that the mean for the entire group of students in the study was equal to 50. In this form, the track means and the groups within track means were as shown in Table 1. The analysis of variance on these means is shown in Table 2. Though the means of Table 1 have been standardized to scale with mean 50 and standard deviation 10, the analysis of variance was performed on the raw data. Since the F-test is invariant and unaffected by change of scale or linear transformation of data, no differences in interpretation can occur.

The F-ratios of the analysis of variance indicated that reliable differences existed only within the Track Two groups. Even though the heterogeneous volunteer group of Track Two did better on the STEP and on the verbal section of the SCAT, the differences are not statistically significant. This is shown by use of the Scheffé confidence interval for the

Table 1. Mean Scores on the Sequential Test of Educational Progress.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	56.7	234
Heterogeneous volunteers	55.4	38
Homogeneous volunteers	56.5	33
Homogeneous non-volunteers	57.0	163
Track Two Students	47.5	268
Heterogeneous volunteers	50.6	37
Homogeneous volunteers	47.6	46
Homogeneous non-volunteers	46.9	185
Track Three Students	39.1	83
Heterogeneous volunteers	41.3	30
Homogeneous volunteers	41.3	6
Homogeneous non-volunteers	37.5	47
Total	50.0	585

Table 2. Analysis of Variance Table for the Sequential Test of Educational Progress.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	50881.82	2	25440.91	
Within Tracks	85022.87	582		
Between Groups-Track One	174.92	2	87.46	.61
Between Groups-Track Two	1003.72	2	501.86	3.47*
Between Groups-Track Three	352.84	2	352.84	2.45
Within Groups Error (Residual)	83138.55	576	144.34**	
Total	135904.69	584	232.71	

*Significant at $\alpha = .05$ $F_{2,576}(.95) = 3.00$

**Mean square within on the mean of 50 and standard deviation of 10 scale is given by MSW = $\frac{144.34}{232.71}(100) = 62.02$.

corresponding comparisons. For this sort of comparison, one examines $\Delta = \hat{\Delta} \pm \underline{S}\ SE_{\hat{\Delta}}$ where $\hat{\Delta} = (\bar{X}_{\text{hetero-volunteer}} - \bar{X}_{\text{homo-volunteer}})$

$\underline{S}\ SE_{\hat{\Delta}} = \sqrt{2F_{2,576}\left(\frac{MSW}{N_1} + \frac{MSW}{N_2}\right)}$. If $\hat{\Delta} > |\pm\underline{S}\ SE_{\hat{\Delta}}|$, it is concluded that the difference is significant. For these data,

$\hat{\Delta} = 50.6 - 47.6 = 3.0$, and $\underline{S}\ SE_{\hat{\Delta}} = \sqrt{2(3.0)\left(\frac{62.02}{37} + \frac{62.02}{46}\right)} = 4.26$.

Since $\hat{\Delta} = 3.0 < 4.26$, it is concluded that the difference is not significant.

No significant differences were found for any of the "within" groups on the SCAT quantitative section or on the total SCAT score which represents an estimate of total ability. The means for the various experimental groups of the study are shown in Tables 3, 5, and 7, while the corresponding analyses of variance are summarized in Tables 4, 6, and 8.

Thus, it can be concluded that with respect to quantitative ability the students within the experimental groups for each track are equivalent. In this sense, volunteering for assignment to a heterogeneous class was not based on quantitative ability, as measured by the SCAT.

While the difference in verbal scores between the heterogeneous and homogeneous volunteers in Track Two is given by $\hat{\Delta} = 50.1 - 46.0 = 4.1$ standard score units, it must be looked upon as a random or chance deviation since both groups of students were generated at random from the one pool of volunteer students, for which the mean score of 48.0 is identical to the mean score of the non-volunteers. Also, it is not sig-

Table 3. Mean Scores on the School and College Ability Test Verbal Section.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	57.4	212
Heterogeneous Volunteers	56.0	38
Homogeneous volunteers	57.3	28
Homogeneous non-volunteers	57.8	146
Track Two Students	48.0	229
Heterogeneous volunteers	50.1	32
Homogeneous volunteers	46.0	42
Homogeneous non-volunteers	48.0	155
Track Three Students	37.9	91
Heterogeneous volunteers	39.7	28
Homogeneous volunteers	35.1	7
Homogeneous non-volunteers	37.3	56
Total	50.0	532

Table 4. Analysis of Variance Table for the School and College Ability Test Verbal Section.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	64226.33	2	32113.16	
Within Tracks	67895.84	529		
Between Groups-Track One	224.93	2	112.47	.88
Between Groups-Track Two	780.32	2	390.16	3.07*
Between Groups-Track Three	415.84	2	207.92	1.64
Within Groups Error (Residual)	66474.75	523	127.10**	
Total	132122.17	531	248.88	

*Significant at $\alpha = .05$ $F_{2,523}(.95) = 3.00$

**Mean square within on the mean 50 and standard deviation 10 scale is given by $MSW = \frac{127.10}{248.88}(100) = 51.07$.

Table 5. Mean Scores on the School and College Ability Test Quantitative Section.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	55.7	212
Heterogeneous volunteers	54.9	38
Homogeneous volunteers	56.6	28
Homogeneous non-volunteers	55.7	146
Track Two Students	49.1	228
Heterogeneous volunteers	50.0	32
Homogeneous volunteers	48.3	42
Homogeneous non-volunteers	49.2	154
Track Three Students	41.9	90
Heterogeneous volunteers	42.5	28
Homogeneous volunteers	40.3	7
Homogeneous non-volunteers	41.8	55
Total	50.0	530

Table 6. Analysis of Variance Table for the School and College Ability Test Quantitative Section.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	53311.28	2	26655.64	
Within Tracks	167613.35	530		
Between Groups-Track One	196.14	2	98.07	.31
Between Groups-Track Two	233.35	2	116.67	.37
Between Groups-Track Three	119.02	2	59.51	.19
Within Groups Error (Residual)	167064.84	524	318.83**	
Total	220924.63	532	415.27	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{318.83}{415.27}(100) = 76.78$.

Table 7. Mean Scores on the School and College Ability Test Total Score.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	57.6	212
Heterogeneous volunteers	55.8	38
Homogeneous volunteers	57.7	28
Homogeneous non-volunteers	58.1	146
Track Two Students	47.7	229
Heterogeneous volunteers	48.7	32
Homogeneous volunteers	46.2	42
Homogeneous non-volunteers	47.8	155
Track Three Students	38.0	91
Heterogeneous volunteers	39.5	28
Homogeneous volunteers	35.5	7
Homogeneous non-volunteers	37.5	56
Total	50.0	532

Table 8. Analysis of Variance Table for the School and College Ability Test Total Score.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	47287.10	2	23643.55	
Within Tracks	46590.43	529		
Between Groups-Track One	280.91	2	140.46	1.61
Between Groups-Track Two	348.96	2	174.48	1.99
Between Groups-Track Three	204.69	2	102.34	1.17
Within Groups Error (Residual)	45755.86	523	87.49**	
Total	93877.52	531	176.79	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{87.49}{176.79}(100) = 49.49$.

nificant by Scheffé's Theorem since $\underline{S}\ SE_{\hat{\Delta}} = 4.11$.

Performance in School Following the First Semester of the 1967-1968 School Year

The overall grade point averages at the end of the first semester indicated that there were no differences between the groups within each track. It would seem that whatever differences may have existed within the Track Two groups, their overall performance as rated by their teachers resulted in no differences. The mean grade point averages, standardized on the total group are as shown in Table 9. The analysis of variance table for these data is shown in Table 10.

Thus, with respect to grade point average, the groups within a track are equivalent following one semester of training. In this sense, heterogeneous grouping in social studies does not affect overall school performance.

Performance in Social Studies and School Following the Second Semester of the 1967-1968 School Year

During the second semester, the Cooperative Social Studies Test was administered to all students in social studies. This test was used to indicate possible differences in achievement that might result from the formation of the experimental classes. The teachers involved indicated that part of this test could be assumed to tap elements of the curriculum which were taught, but that part of the test contained questions about materials which were not taught but might be peripherally learned. For this reason, the items on the test were divided into two sets. Set One contained the questions that seemed, to the teachers, to be testing what the students should have

Table 9. Mean Grade Point Averages Over All Classes at the End of the First Semester.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	56.3	212
Heterogeneous volunteers	54.1	38
Homogeneous volunteers	56.9	28
Homogeneous non-volunteers	56.8	146
Track Two Students	47.9	228
Heterogeneous volunteers	47.5	32
Homogeneous volunteers	46.5	42
Homogeneous non-volunteers	48.4	154
Track Three Students	40.4	90
Heterogeneous volunteers	42.8	28
Homogeneous volunteers	40.1	7
Homogeneous non-volunteers	39.2	55
Total	50.0	530

Table 10. Analysis of Variance Table for Grade Point Averages Over all Classes at the End of the First Semester.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	101.99	2	50.99	
Within Tracks	201.78	527		
Between Groups-Track One	1.32	2	.66	1.73
Between Groups-Track Two	.69	2	.35	.92
Between Groups-Track Three	1.41	2	.70	1.84
Within Groups Error (Residual)	198.36	521	.38**	
Total	303.76	529	.57	

**Mean square within on the mean of 50 and standard deviation of 10 scale is given by MSW = $\frac{.38}{.57}(100) = 66.67$

learned; Set Two contained the questions which seemed, to the teachers, to be testing an area which was not directly taught, but which might have been peripherally learned.

The resulting statistics indicate that the Track One students in the heterogeneous classes did less well on both versions of this test, and that the Track Two and Track Three students in the heterogeneous classes did as well, but not better, than the other students in their respective tracks.

The mean scores of the two tests, standardized as before, are shown in Tables 11 and 13. The corresponding analysis of variance tables are shown in Tables 12 and 14.

For the relevant test items the mean differences between the heterogeneous and homogeneous volunteers in Track One is given by $\hat{\Delta} = 52.7 - 56.4 = -3.7$ standard units with $\underline{S}\ SE_{\hat{\Delta}} = 4.74$ so that the difference is not statistically significant. However, if one compares the heterogeneous students to the homogeneous students, $\hat{\Delta} = 52.7 - 57.3 = -4.6$ with $\underline{S}\ SE_{\hat{\Delta}} = 3.44$ indicating that a significant difference exists. With respect to the non-relevant or peripherally learned items, the corresponding mean difference is given by $\hat{\Delta} = 52.0 - 56.0 = -4.0$ with $\underline{S}\ SE_{\hat{\Delta}} = 5.23$. Again, this difference is not significant. For the comparison of heterogeneous students versus homogeneous students the difference is given by $\hat{\Delta} = 52.0 - 56.0 = -4.0$ with $\underline{S}\ SE_{\hat{\Delta}} = 3.80$ indicating that an obvious reason for the significant F-ratio is not apparent.

However, if these differences are considered important, then it must be concluded that top ability students did not perform in heterogeneous classes as well as they performed in

Table 11. Mean Scores on the Cooperative Social Studies Test Version One.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	56.4	213
Heterogeneous volunteers	52.7	40
Homogeneous volunteers	56.4	30
Homogeneous non-volunteers	57.4	143
Track Two Students	48.1	229
Heterogeneous volunteers	46.5	33
Homogeneous volunteers	47.2	41
Homogeneous non-volunteers	48.6	155
Track Three Students	40.0	93
Heterogeneous volunteers	40.2	29
Homogeneous volunteers	43.0	6
Homogeneous non-volunteers	39.7	58
Total	50.0	535

Table 12. Analysis of Variance Table for Cooperative Social Studies Test Version One.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	11997.86	2	5998.93	
Within Tracks	21921.93	532		
Between Groups-Track One	364.69	2	182.34	4.47*
Between Groups-Track Two	100.03	2	50.02	1.23
Between Groups-Track Three	38.84	2	19.42	.48
Within Groups Error (Residual)	21418.37	526	40.72**	
Total	33919.79	534	63.52	

*Significant at $\alpha = .05$ $\quad F_{2,526}(.95) = 3.00$

**Mean square within on the mean 50 and standard deviation 10 scale is given by $MSW = \frac{40.72}{63.52}(100) = 64.11$.

Table 13. Mean Scores on the Cooperative Social Studies Test Version Two.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	55.3	213
Heterogeneous volunteers	52.0	40
Homogeneous volunteers	56.0	30
Homogeneous non-volunteers	56.0	143
Track Two Students	48.0	229
Heterogeneous volunteers	46.9	33
Homogeneous volunteers	47.1	41
Homogeneous non-volunteers	48.4	155
Track Three Students	42.9	93
Heterogeneous volunteers	44.3	29
Homogeneous volunteers	42.3	6
Homogeneous non-volunteers	42.3	58
Total	50.0	535

Table 14. Analysis of Variance Table for the Cooperative Social Studies Test Version Two.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	1504.13	2	752.06	
Within Tracks	5506.22	532		
Between Groups-Track One	67.82	2	33.91	3.29*
Between Groups-Track Two	13.35	2	6.67	.65
Between Groups-Track Three	10.95	2	5.48	.53
Within Groups Error (Residual)	5414.10	526	10.29**	
Total	7010.35	534	13.13	

*Significant at $\alpha = .05$ $\quad F_{2,526}(.95) = 3.00$

**Mean square within on the mean 50 and standard deviation 10 scale is given by:

$$MSW = \frac{10.29}{13.13}(100) = 78.37.$$

homogeneous classes. In case this decision must be made with reservations since the contrast of heterogeneous versus homogeneous students in the volunteer group is not significant. On the other hand, students of Tracks Two and Three perform equally well in heterogeneous or homogeneous classes.

These results correspond with teacher's comments concerning their inability to motivate the Track One students placed in the heterogeneous classes.

On the peripheral items, the Track One heterogeneous students scored four points below the Track One homogeneous students. Thus, again, it appears that Track One students trained in heterogeneous classes do not perform as well as expected, but again, it must be recalled that the comparison of the heterogeneous and homogeneous students in the volunteer group is not significant. For the Track Two and Three students no differences are found between students trained in heterogeneous and homogeneous classes.

A second criterion measure of social studies achievement was devised using the test on the constitution required for the eighth grade. Previous to the teaching of the unit specifically preparing the students for the test, a selection of questions supplied by the textbook publishers was given to all students in the study to see if there were initial differences between the groups within the tracks. Subsequent to the teaching of the unit on the constitution, a test, again using questions supplied by the textbook publishers, was given to indicate differences in achievement on the unit.

The results indicated that on the pretest there were

no significant differences within the tracks. The results on the post test indicated that the Track Three volunteers in the heterogeneous classes did get higher scores than the other Track Three students. This may indicate that heterogeneous classes were taught with greater reliance on the textbook as a guide, or alternatively, more concrete questions on facts from the text tapped differences in the Track Three students which the Cooperative Social Studies Test did not. The Track One and Track Two students did as well as, but not better than, the other students who were taught in the homogeneous classes.

The means, standardized as before, are reported in Tables 15 and 17. The corresponding analyses of variance are reported in Tables 16 and 18. If the heterogeneous students of Track Three are compared to the corresponding homogeneous students, it is seen that the mean difference is given by $\hat{\Delta} = 42.9 - 38.0 = 4.9$, or almost one-half of a standard deviation difference. With $\underline{S}\ SE_{\hat{\Delta}} = 4.90$, it is seen that the difference is significant. Thus, on this part of the test, hypothesis three of the study is supported. Track Three students placed in the heterogeneous classes did better than the Track Three students placed in the homogeneous classes.

The overall grade point average for the end of the school year indicated that there were no differences between the groups within the various tracks. The mean scores are reported in Table 19, while the analysis of variance is reported in Table 20.

Attitudes Toward Classwork and Other Students in Social Studies

A series of attitude questionnaires was administered

Table 15. Mean Scores on the Pre Test on the U. S. Constitution Teaching Unit.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	55.4	195
Heterogeneous volunteers	54.1	36
Homogeneous volunteers	54.9	25
Homogeneous non-volunteers	55.9	134
Track Two Students	47.1	200
Heterogeneous volunteers	48.8	27
Homogeneous volunteers	48.7	33
Homogeneous non-volunteers	46.5	140
Track Three Students	43.4	74
Heterogeneous volunteers	45.0	26
Homogeneous volunteers	46.5	6
Homogeneous non-volunteers	42.5	42
Total	50.0	469

Table 16. Analysis of Variance Table for the Pre Test on the U. S. Constitution Teaching Unit.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	6082.87	2	3041.43	
Within Tracks	20811.49	466		
Between Groups-Track One	54.09	2	27.04	.61
Between Groups-Track Two	126.54	2	63.27	1.41
Between Groups-Track Three	60.21	2	30.10	.67
Within Groups Error (Residual)	20570.65	460	44.71**	
Total	26894.36	468	57.47	

**Mean square witin on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{44.72}{57.47}(100) = 77.81$.

Table 17. Mean Scores on the Post Test on the U. S. Constitution Teaching Unit.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	55.8	195
Heterogeneous volunteers	54.1	36
Homogeneous volunteers	55.5	25
Homogeneous non-volunteers	56.3	134
Track Two Students	48.2	200
Heterogeneous volunteers	47.6	27
Homogeneous volunteers	45.1	33
Homogeneous non-volunteers	48.8	140
Track Three Students	39.7	74
Heterogeneous volunteers	42.9	26
Homogeneous volunteers	40.3	6
Homogeneous non-volunteers	37.7	42
Total	50.0	469

Table 18. Analysis of Variance Table for the Post Test on the U. S. Constitution Teaching Unit.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	15443.41	2	7721.70	
Within Tracks	33031.65	466		
Between Groups-Track One	134.77	2	67.38	.96
Between Groups-Track Two	206.17	2	103.09	1.47
Between Groups-Track Three	444.09	2	222.04	3.17*
Within Groups Error (Residual)	32246.62	460	70.10**	
Total	48475.06	468	103.58	

*Significant at $\alpha = .05$ $F_{2,460}(.95) = 3.00$

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{70.10}{103.58}(100) = 67.68$.

Table 19. Mean Grade Point Averages Over all Classes at the End of the Second Semester.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	55.9	215
Heterogeneous volunteers	53.7	39
Homogeneous volunteers	56.3	29
Homogeneous non-volunteers	56.5	147
Track Two Students	47.8	231
Heterogeneous volunteers	47.6	33
Homogeneous volunteers	45.9	43
Homogeneous non-volunteers	48.3	155
Track Three Students	41.9	95
Heterogeneous volunteers	43.6	30
Homogeneous volunteers	46.0	7
Homogeneous non-volunteers	40.6	58
Total	50.0	541

Table 20. Analysis of Variance Table for the Grade Point Averages Over All Classes at the End of the Second Semester.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	96.73	2	48.37	
Within Tracks	250.99	538		
Between Groups-Track One	1.57	2	.75	1.71
Between Groups-Track Two	1.26	2	.63	1.43
Between Groups-Track Three	1.99	2	.99	2.25
Within Groups Error (Residual)	246.23	532	.44**	
Total	347.72	540	.64	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{.44}{.65}(100) = 67.69$.

during the second semester. The exact questionnaire employed is reproduced in Appendix A. The items in this questionnaire were aimed at assessing attitudes toward self, school, social studies classes, and other relevant areas. The responses to these questionnaires were submitted to a factor analysis, whereupon those questions which did not appear, on the basis of the preliminary run, as discriminating questions were eliminated from the battery; the remaining questions were factor analyzed a second time. The measures used for comparison were factor scores based upon the rotated factors from the second analysis. These scores were standardized across the total group so that the average is equal to 50, with a standard deviation of 10.

The adjective pairs that were used to define the dimension of "self" are reported in Table 21 along with their factor weights. As can be seen, Factor One is a self-evaluative factor. Students who score high on this dimension report themselves as being pleasant, happy, valuable, beautiful, etc., whereas students with poor self-images see themselves as unpleasant, sad, worthless, ugly, etc.

In the evaluation of "self," reliable differences existed between the tracks and within Track Three. While various interpretations may be offered to explain the results of the analysis of variance on the factor scores, a description of the differences might be summarized as follows: On the whole, Track One students are more self-critical, perhaps because they are more sensitive to differences between their conception of "ideal" behavior and their own behavior and because the potential and real activities of their peers are of

Table 21. Adjective Pairs That Define Factor One: How I Feel About Myself.

Item Number	Adjective Pairs	Factor Coefficient
25	pleasant - unpleasant	.70
28	happy - sad	.65
23	valuable - worthless	.64
24	beautiful - ugly	.63
34	fragrant - foul	.62
41	hopeful - hopeless	.61
33	nice - awful	.59
44	fast - slow	.54
46	top - bottom	.53
35	healthy - sick	.53
11	brave - cowardly	.53
37	active - passive	.52
6	free - bound	.51
8	strong - weak	.51

Prcportion of variance explained by Factor One is equal to 21 percent.

a higher standard. The more interesting result of the analysis occurred in the three groups comprising the Track Three students. Those volunteers who were assigned to the experimental classes exhibited the same tendency toward self-criticism. This may represent the acquisition of a more realistic interpretation of their own behavior or it may be simply the acquisition of new attitudes from their heterogeneous group since this measure was administered in the context of the social studies classes.

The means of the groups are shown in Table 22 with the corresponding analysis of variance table shown in Table 23. For the Track Three students the mean difference for heterogeneous versus homogeneous students is given by $\hat{\Delta} = 47.3 - 55.8 = -8.5$ standard units with $\underline{S}\ SE_{\hat{\Delta}} = 6.97$. If this poorer self-image represents a true or more honest appraisal by these students, and since it is quite similar to the Track One mean of 48.3, then it would appear that heterogeneous grouping is indeed a valuable adjunct to determining self-worth. Students who are complacent and satisfied will not try to improve their status and standing whereas a student who believes there is room for improvement will try harder and seek to improve. Such an improvement was noted for the Constitution post test. In this sense, a more honest self-appraisal with a reduced self-image score was accompanied with increased school learning on the Constitution Test.

It has been stated that minority students generally have poor self-images which are further damaged by being placed in low ability tracks. Since information was available for the race of the students, it was decided to make an investigation

Table 22. Mean Scores for the Evaluation of Self.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	48.3	171
Heterogeneous volunteers	47.7	34
Homogeneous volunteers	45.3	25
Homogeneous non-volunteers	49.2	112
Track Two Students	50.9	172
Heterogeneous volunteers	51.1	26
Homogeneous volunteers	49.6	29
Homogeneous non-volunteers	51.2	117
Track Three Students	54.2	56
Heterogeneous volunteers	47.3	17
Homogeneous volunteers	58.0	4
Homogeneous non-volunteers	55.5	35
Total	50.0	399

Table 23. Analysis of Variance Table for Evaluation of Self.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	561.91	2	280.95	
Within Tracks	17887.29	396		
Between Groups-Track One	146.52	2	73.26	1.65
Between Groups-Track Two	29.04	2	14.52	.33
Between Groups-Track Three	403.31	2	201.66	4.54*
Within Groups Error (Residual)	17308.42	390	44.38**	
Total	18449.20	398	46.35	

*Significant at $\alpha = .05$ $F_{2,390}(.95) = 3.00$

**Mean square within on the mean of 50 and standard deviation of 10 scale is given by $MSW = \frac{44.38}{46.35}(100) = 95.75$.

of this commonly heard statement. For the most part, the responses made by the students, independent of their class assignment does not support this hypothesis. For the Negro students, the mean standardized scores across the three tracks are given by 55.4, 53.6, and 54.1. Thus, Negro students in all tracks have positive self-images as measured by the items of the adjective check list. Furthermore, the positiveness of self-image does not decline as class ability decreases. Thus, high achieving Negro students and low achieving Negro students report positive self-images. For the Caucasian students, the mean standardized scores across the three tracks are given by 48.0, 47.3, and 57.5. In this case, the lowest achieving Caucasians express a very strong positive self-image, while the Caucasians of Tracks One and Two report the poorest.

It should be noted that these results do not agree completely with those reported by Coleman (p. 281). For his investigation, "Three questions were used at the ... 9th grade level[s] to obtain an indication of the child's self-concept. These are:

(1) How bright do you think you are in comparison with the other students in your grade?

(2) Agree or disagree: I sometimes feel that I just can't learn.

(3) Agree or disagree: I would do better in school work if teachers didn't go so fast.

In general, the responses to these questions do not indicate differences between Negroes and whites, ..." According to Coleman, "Negroes and whites show similar levels of response

to these items, ..." He then continues, "It is puzzling to some analysts that the Negro children report levels of self-esteem as high as whites when there is so much in their social environment to reduce the self-esteem of a Negro, and those analysts conjecture that these responses may not mean what their face value suggests." It should be noted that Marascuilo (1969) has reported strong positive self-images for a reading study conducted by the Berkeley Unified School District under the auspices of California State Bill 28. As reported in the Summary of the Report (15 September 1969):

"...it is conceivable that student's attitudes are not as negative as some teachers and administrators believe. For the most part, student attitudes on the pre test were quite positive. While they may represent a false or invalid measure of attitude, they were of such a nature that they would be difficult to improve or if they did improve, to measure the improvement." It is quite possible that these decisions are applicable to this study on the effects of homogeneous versus heterogeneous groupings.

From the factor analysis conducted on the 50 adjective pairs that were used to measure attitude toward self, a second factor was identified. The adjective pairs that define this second factor are shown in Table 24. Students who rank themselves high on this dimension report themselves as boyish, hard, manly, rugged, rough, and bully, while students at the opposite pole report themselves as girlish, soft, unmanly, delicate, smooth, and gentle. This suggests that this second factor is related to sexual identity--an important psychological

Table 24. Adjective Pairs That Define Factor II: Sexual Identity.

Item Number	Adjective Pairs			Factor Coefficient
49	boyish	-	girlish	.76
22	hard	-	soft	.74
40	manly	-	womanly	.73
20	rugged	-	delicate	.69
13	rough	-	smooth	.60
14	bully	-	gentle	.54

Proportion of variance explained by Factor II is equal to 12 percent.

characteristic at this particular age group. The mean scores for the nine experimental groups are shown in Table 25, while the corresponding analysis of variance is reported in Table 26. For this variable, there are no statistically significant differences between the tracks nor within the tracks.

A second factor analysis was performed on the adjective pairs that were used to define the way students felt about the other students in the school. The results of the analysis are shown in Tables 27 and 30. On the first factor, students with positive feelings about the other students report them as being nice, fragrant, clean, sweet, fair, etc., while students with unfavorable feelings report their fellow classmates as awful, foul, dirty, sour, unfair, etc. This factor is considerably unlike Factor I related to self. While Factor I towards self is an evaluate factor concerning worthyness and value of a person, Factor I towards the other students is a factor describing how students behave as perceived by students making the evaluation. Students with favorable attitudes on this dimension report their peers as nice, clean, and fair, while at the opposite pole students are reported as awful, dirty, and unfair. The mean scores for the nine study groups are presented in Table 28, while the analysis of variance is reported in Table 29. As can be seen, none of the mean differences are statistically significant, whether they are made between the tracks or within the tracks.

The second factor for this set of items appears to be sexual in nature, but involves a component of activity as suggested by the appearance of the adjective pair (talkative--not

Table 25. Mean Scores for Factor II: Sexual Identity.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	49.9	171
Heterogeneous volunteers	49.4	34
Homogeneous Volunteers	48.8	25
Homogeneous non-volunteers	50.2	112
Track Two Students	49.5	172
Heterogeneous volunteers	47.8	26
Homogeneous volunteers	47.1	29
Homogeneous non-volunteers	50.5	117
Track Three Students	51.8	56
Heterogeneous volunteers	52.6	17
Homogeneous volunteers	46.0	4
Homogeneous non-volunteers	48.0	35
Total	50.0	399

Table 26. Analysis of Variance Table for Factor II: Sexual Identity.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	33.91	2	16.96	
Within Tracks	6079.06	396		
Between Groups-Track One	7.75	2	3.87	.25
Between Groups-Track Two	56.10	2	28.05	1.82
Between Groups-Track Three	22.66	2	11.33	.73
Within Groups Error (Residual)	5992.55	390	15.37**	
Total	6112.97	398	15.32	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{15.37}{15.32}(100) = 100.29$.

Table 27. Adjective Pairs that Define Factor I: How I Feel About the Other Kids.

Item Number	Adjective Pairs			Factor Coefficient
33	nice	-	awful	.77
34	fragrant	-	foul	.72
19	clean	-	dirty	.71
17	sweet	-	sour	.71
36	fair	-	unfair	.70
27	sweet	-	bitter	.70
25	pleasant	-	unpleasant	.70
42	smart	-	dumb	.70
21	tasteful	-	distasteful	.69
41	hopeful	-	hopeless	.69
35	healthy	-	sick	.68
16	good	-	bad	.67
23	valuable	-	worthless	.64
24	beautiful	-	ugly	.60
31	success	-	failure	.60
2	honest	-	dishonest	.59
28	happy	-	sad	.58
47	interested	-	bored	.57
39	rich	-	poor	.56
18	equal	-	unequal	.55
12	human	-	animal	.54
46	top	-	bottom	.54
30	peaceful	-	ferocious	.53
14	gentle	-	bully	.52

Proportion of variance explained by Factor I equal to 30 percent.

Table 28. Mean Scores for Factor I: How I Feel About the Other Kids.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	50.5	132
Heterogeneous volunteers	49.8	27
Homogeneous volunteers	49.9	23
Homogeneous non-volunteers	50.9	82
Track Two Students	49.7	125
Heterogeneous volunteers	49.7	22
Homogeneous volunteers	45.9	20
Homogeneous non-volunteers	50.6	83
Track Three Students	49.4	46
Heterogeneous volunteers	50.7	15
Homogeneous volunteers	45.9	4
Homogeneous non-volunteers	49.2	27
Total	50.0	303

Table 29. Analysis of Variance Table for Factor I: How I Feel About the Other Kids.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	72.44	2	36.22	
Within Tracks	39368.57	300		
Between Groups-Track One	44.11	2	22.06	.17
Between Groups-Track Two	475.74	2	237.87	1.80
Between Groups-Track Three	95.68	2	47.84	.36
Within Groups Error (Residual)	38753.04	294	131.81**	
Total	39441.01	302	130.60	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{131.81}{130.60}(100) = 100.93$.

Table 30. Adjective Pairs that Define Factor II: How I Feel About the Other Kids.

Item Number	Adjective Pairs	Factor Coefficient
22	hard - soft	.61
20	rugged - delicate	.54
40	manly - womanly	.52
29	talkative - not talkative	.52

Proportion of variance explained by Factor I is equal to 37 percent

talkative). The mean scores for the nine study groups are shown in Table 31, while the analysis of variance is shown in Table 32. Since the F value is quite small for the between tracks square of variance, it follows that none of the mean differences are statistically significant. This is also true for the mean differences in Track Two. However, the observed differences for Track One and Track Three are larger than would be expected by chance alone. It appears that the volunteers of Track One view their peers as more rugged, manly, and talkative. For the Track Three students it appears that the homogeneous volunteers find their peers as more feminine than do the other groups of students while the heterogeneous volunteers find them more masculine. It is not easy to produce an explanation for this finding. It is quite possible that students placed in the heterogeneous classes did more talking or may have appeared to be more disruptive of class activities. However, there is no way to verify this.

Sets of questions which were designed to tap attitudes towards various aspects of the social studies classes for the previous and for the experimental year were also evaluated. This was done by means of a factor analysis followed by a set of univariate analyses of variance.

The items that define attitude towards classwork and class assignments in social studies last year are summarized in Table 33. As can be seen, the six items that define this characteristic go logically together. Question 9 appears with a negative coefficient suggesting that if the response choices had been inverted to correspond to the five other questions, it

Table 31. Mean Scores for Factor II: Sex-Role Characteristics of Others in Class.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	49.7	132
Heterogeneous volunteers	53.6	27
Homogeneous volunteers	50.7	23
Homogeneous non-volunteers	48.1	82
Track Two Students	49.7	125
Heterogeneous volunteers	48.3	22
Homogeneous volunteers	51.3	20
Homogeneous non-volunteers	49.7	83
Track Three Students	51.6	46
Heterogeneous volunteers	56.0	15
Homogeneous volunteers	41.8	4
Homogeneous non-volunteers	50.6	27
Total	50.0	303

Table 32. Analysis of Variance Table for Factor II: Sex-Role Characteristics of Others in Class.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	10.42	2	5.21	
Within Tracks	2170.03	300		
Between Groups-Track One	46.21	2	23.11	3.29*
Between Groups-Track Two	6.78	2	3.39	.48
Between Groups-Track Three	50.87	2	25.44	3.62*
Within Groups Error (Residual)	2066.17	294	7.03**	
Total	2180.45	302	7.22	

*Significant at $\alpha = .05$ $\quad F_{2,294}(.95) = 3.00$

**Mean square within on the mean of 50 and standard deviation of 10 scale is given by MSW = $\frac{7.03}{7.22}(100) = 97.36$.

Table 33. Items that Define Attitudes Towards Classwork and Assignments in Social Studies Last Year.

Item Number	Question	Factor Coefficient
4	Last year in World History and Geography I liked to come to class (circle your answer) ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME	.78
13	Last year in World History I was happy to do the class assignments MOST OF THE TIME MUCH OF THE TIME ONCE IN A WHILE HARDLY EVER	.75
1	How well did you like the classwork in World History and Geography last year? VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL	.73
9	Last year in World History and Geography I enjoyed doing extra work for my teacher. HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME	-.69
7	How much did you learn about World History and Geography last year? QUITE A LOT A LOT A LITTLE NOTHING	.61
8	How well did you like your teacher in World History and Geography last year? VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL	.61

Proportion of variance accounted for by Factor I: 20 percent.

would have appeared in the factor analysis with a positive coefficient. The means for the nine study groups are shown in Table 34 and the analysis of variance is shown in Table 35. None of the differences between the tracks are significant nor are any of the differences within a track significant.

The items that define attitude towards classmates in social studies last year are summarized in Table 36. The mean scores for the heterogeneous and homogeneous groups in the various tracks are reported in Table 37 and the corresponding analysis of variance is reported in Table 38. In this case there are large differences between the tracks but no differences within the tracks. Students in Track One tended to report that the kids in World History and Geography were as smart as them, they hardly ever felt impatient with the slow kids, they felt at ease with the other kids, and they respected the other kids as individuals. On the other hand, Track Three students tended to report the opposite about the kids that were in their World History and Geography classes.

The responses to the questions related to the previous school year indicated that mean differences existed only between the tracks. Track One students, regardless of their volunteer status, had a more positive attitude toward their classmates during the previous year. No differences existed at all between or within the groups regarding the attitudes toward classwork and assignments.

For the questions aimed at the social studies classes for the first experimental year, the Track One students exhibited similar results on questions about their classmates. On ques-

Table 34. Mean Scores for Attitudes Toward Assignments and Classwork in Social Studies Last Year.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	50.3	180
Heterogeneous volunteers	49.3	35
Homogeneous volunteers	47.3	28
Homogeneous non-volunteers	51.3	117
Track Two Students	49.4	199
Heterogeneous volunteers	50.7	30
Homogeneous volunteers	51.4	30
Homogeneous non-volunteers	48.6	139
Track Three Students	51.2	56
Heterogeneous volunteers	50.1	21
Homogeneous volunteers	54.5	6
Homogeneous non-volunteers	51.3	29
Total	50.0	435

Table 35. Analysis of Variance Table for Attitudes Toward Assignments and Classwork in Social Studies Last Year.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	33.98	2	16.99	
Within Tracks	7989.18	432		
Between Groups-Track One	74.41	2	37.21	2.01
Between Groups-Track Two	49.45	2	24.73	1.34
Between Groups-Track Three	16.64	2	16.64	.90
Within Groups Error (Residual)	7848.68	426	18.42**	
Total	8023.16	434	18.49	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{18.42}{18.49}(100) = 99.62$.

Table 36. Items that Define Attitudes Towards Classmates in Social Studies Last Year.

Item Number	Question	Factor Coefficient
2	Last year the kids in World History and Geography were as smart as I am. ALL OF THEM MOST OF THEM FEW OF THEM NONE OF THEM	.56
6	Last year I felt impatient with some of the slow kids in my World History and Geography class. HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME	.54
19	Last year I felt at ease with the other kids in my World History and Geography class. ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME	.50
23	Last year among the other students in my World History and Geography class I respected as individuals ALL OF THEM MOST OF THEM FEW OF THEM NONE OF THEM	.44
21	Last year there were times when my behavior towards the other kids in World History and Geography was quite different from the way I felt inside. ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME	-.43

Proportion of variance accounted for by Factor II: 9 percent.

Table 37. Mean Scores for Attitudes Towards Others in the Class in Social Studies Last Year.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	53.1	180
Heterogeneous volunteers	51.8	35
Homogeneous volunteers	51.0	28
Homogeneous non-volunteers	54.0	117
Track Two Students	48.6	200
Heterogeneous volunteers	48.0	30
Homogeneous volunteers	50.0	31
Homogeneous non-volunteers	48.4	139
Track Three Students	45.1	55
Heterogeneous volunteers	45.3	21
Homogeneous volunteers	48.9	5
Homogeneous non-volunteers	44.2	29
Total	50.0	435

Table 38. Analysis of Variance Table for Attitudes Towards Others in the Class in Social Studies Last Year.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	242.97	2	121.49	
Within Tracks	2806.91	432		
Between Groups-Track One	18.61	2	9.31	1.43
Between Groups-Track Two	5.75	2	2.88	.44
Between Groups-Track Three	6.65	2	3.32	.51
Within Groups Error (Residual)	2775.90	426	6.51**	
Total	3049.88	434	7.03	

**Mean squares within on the mean 50 and standard deviation 10 scale is given by:

$$MSW = \frac{6.52}{7.03}(100) = 92.75.$$

tions aimed at attitudes towards classwork and assignments this year, the Track One heterogeneous group liked this aspect less than the other Track One groups liked theirs. In Track Two, both the selected and non-selected volunteer groups liked their classwork and assignments less than the non-volunteer group. In Track Three, the heterogeneous group liked their classwork, etc., less well than either of the other two Track Three groups.

The items that define attitudes towards classwork and assignments in social studies during the 1967-1968 experimental year are shown in Table 39. If these items are compared to the items of Table 33, it is seen that they are exactly the same. The mean scores for the nine study groups are shown in Table 40 and the corresponding analysis of variance is reported in Table 41. In this case, each within track source of variance is statistically significant. In each case, it is the heterogeneous volunteers that report the least satisfaction with the assignments and classwork. This dissatisfaction exists for students who would normally have been assigned to Tracks One, Two, or Three so that it cannot be concluded that the dissatisfaction is expressed solely by the top ability or low ability students. The mean differences for the three tracks of students are given by:

$$\hat{\Delta}_1 = 51.5 - 46.2 = 5.3 \text{ with } \underline{S}\ SE_{\hat{\Delta}} = 4.51$$

$$\hat{\Delta}_2 = 50.7 - 46.3 = 4.4 \text{ with } \underline{S}\ SE_{\hat{\Delta}} = 4.73$$

$$\hat{\Delta}_3 = 52.9 - 44.7 = 8.2 \text{ with } \underline{S}\ SE_{\hat{\Delta}} = 6.37$$

suggesting that the dissatisfaction increases as the track number increases. This is somewhat surprising when it is re-

Table 39. Items that Define Attitude Towards Classword and Assignments in Social Studies This Year.

Item Number	Question	Factor Coefficient
27	This year in U. S. History I like to come to class ALL THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME	.78
31	How well do you like the teacher in U. S. History this year? VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL	.75
36	This year in U. S. History I am happy to do the class assignments. MOST OF THE TIME MUCH OF THE TIME ONCE IN A WHILE HARDLY EVER	.75
24	How well do you like the classwork in U. S. History this year? VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL	.72
32	This year in U. S. History I enjoy doing the extra work for my teacher. HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME	-.70
30	How much are you learning about U. S. History this year? QUITE A LOT A LOT A LITTLE NOTHING	.66

Proportion of variance accounted for by Factor I: 20 percent.

Table 40. Mean Scores for Attitudes Toward Assignments and Classwork in Social Studies This Year.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	50.4	180
Heterogeneous volunteers	46.2	35
Homogeneous volunteers	51.0	28
Homogeneous non-volunteers	51.6	117
Track Two Students	49.5	191
Heterogeneous volunteers	46.3	30
Homogeneous volunteers	46.8	31
Homogeneous non-volunteers	50.9	130
Track Three Students	50.3	64
Heterogeneous volunteers	44.7	21
Homogeneous volunteers	55.0	5
Homogeneous non-volunteers	52.7	38
Total	50.0	435

Table 41. Analysis of Variance Table for Attitudes Toward Assignments and Classwork in Social Studies This Year.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	15.63	2	7.82	
Within Tracks	7984.33	432		
Between Groups-Track One	147.54	2	73.77	4.19*
Between Groups-Track Two	149.88	2	74.94	4.26*
Between Groups-Track Three	183.40	2	91.70	5.21*
Within Groups Error (Residual)	7503.51	426	17.61**	
Total	7999.96	434	18.43	

*Significant at $\alpha = .05$ $F_{2,426}(.95) = 3.00$

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{17.61}{18.43}(100) = 95.55$.

called that the Track Three students showed considerable learning on the Constitution post test.

It should be noted that part of the dissatisfaction could be related to attitudes towards the teacher, for if the position of this item in Tables 33 and 39 is compared it is seen that this item has a higher ranking on the hierarchy of factor weights at the later year. The remaining items maintain their same relative positions to one another.

The items that define attitudes towards classmates in social studies are shown in Table 42. These items are essentially the same as the items used for the previous year's variable. The mean scores for the various groups are reported in Table 43 and the corresponding analysis of variance is reported in Table 44. None of the differences within the individual tracks are significant. Only the difference between the tracks is statistically significant but this difference parallels that noticed for the corresponding items relating to last year.

Attitudes of the Adults at Home to the School Environment as Perceived by the Students

A series of questions designed to elicit attitudes of the adults at home as projected by the students yielded some differences. The questions were worded so that the students were to answer on the basis of what they knew or believed about the attitudes of their parents or guardians. It was not expected that this would necessarily reflect the true attitudes of the adults in the family.

Track One students as a whole said the adults in their

Table 42. Items That Define Attitudes Towards Classmates in Social Studies This Year.

Item Number	Question	Factor Coefficient
42	This year I feel at ease with the other kids in my U. S. History class. ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME	.63
44	This year there are times when my behavior towards the other kids in U. S. History is quite different from the way I feel inside. ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME	-.60
34	This year I feel impatient with some of the fast kids in my U. S. History class. HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME	.59
46	This year among the other students in my U. S. History class I respect them as individuals ALL OF THEM MOST OF THEM FEW OF THEM NONE OF THEM	.53
29	This year I feel impatient with some of the slow kids in my U. S. History class. HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME	.46

Proportion of variance accounted for by Factor II: 9 percent.

Table 43. Mean Scores for Attitudes Towards Others in the Class in Social Studies This Year.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	52.6	180
Heterogeneous volunteers	50.6	35
Homogeneous volunteers	50.1	28
Homogeneous non-volunteers	54.0	117
Track Two Students	49.4	191
Heterogeneous volunteers	46.8	30
Homogeneous volunteers	47.5	31
Homogeneous non-volunteers	50.5	130
Track Three Students	44.4	64
Heterogeneous volunteers	40.2	21
Homogeneous volunteers	49.0	5
Homogeneous non-volunteers	45.6	38
Total	50.0	435

Table 44. Analysis of Variance Table for Attitudes Towards Others in the Class in Social Studies This Year.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	260.45	2	130.22	
Within Tracks	3136.45	432		
Between Groups-Track One	38.24	2	19.12	2.69
Between Groups-Track Two	32.21	2	16.10	2.26
Between Groups-Track Three	37.38	2	18.69	2.63
Within Groups Error (Residual)	3028.62	426	7.11**	
Total	3396.90	434	7.83	

**Mean square within on the mean 50 and standard deviation 10 scale is given by $MSW = \frac{7.11}{7.83}(100) = 90.80$.

home approved of the behavior of the student body and the educational quality of the school Track Two students reflected a less approving attitude and the Track Three students reflected the least approving attitude. There were no differences on this variable for any of the within track group comparisons.

On the questions regarding grouping or tracking in the school, the Track One volunteers assigned to the heterogeneous classes said their parents disapproved of tracking, the Track One volunteers relegated to the regular classes said their parents disapproved of tracking (but not to the degree of the experimental group), while the Track One non-volunteers said that their parents approved of ability grouping in school. There were no reliable differences between the tracks or within Tracks Two and Three.

The items that define what the adults at home think about school and its educational quality, its ability group tracks, and its administration and teachers are shown in Tables 45, 48, and 51. The corresponding tables of means are shown in Tables 46, 49, and 52. The analysis of variance tables for these three variables are shown in Tables 47, 50, and 53.

Summary Conclusions for the 1967-1968 School Year

The general results for the 1967-1968 school year are not impressive. There were indications, not statistically significant, that the Track One students did not achieve as well under heterogeneous grouping by about .3 to .4 of a standard deviation on the standardized score scale. The teachers involved with the heterogeneous classes reported difficulties in

Table 45. Items that Define What the Adults at Home Think About School and Its Educational Quality.

Item Number	Question	Factor Coefficient
6	The adults at home think this school is below standard. TRUE FALSE	.64
21	The adults at home think I should go to a different school. TRUE FALSE	.63
4	The adults at home think my counselors are not helpful. TRUE FALSE	.48
7	The adults at home think there are too many extra-curricular activities. TRUE FALSE	.47
16	The adults at home think school is a waste of time. TRUE FALSE	.46
23	The adults at home think discipline at this school is weak. TRUE FALSE	.43
8	The adults at home think my teachers are fair. TRUE FALSE	.43
1	The adults at home think I am getting a good education at this school. TRUE FALSE	.42

Proportion of variance accounted for by Factor I: 16 percent.

Table 46. Mean Scores of What the Adults at Home Think About Student Behavior and General Educational Quality.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	51.6	184
Heterogeneous volunteers	51.8	35
Homogeneous volunteers	53.1	28
Homogeneous non-volunteers	51.3	121
Track Two Students	49.2	203
Heterogeneous volunteers	48.0	29
Homogeneous volunteers	50.3	33
Homogeneous non-volunteers	49.2	141
Track Three Students	46.9	57
Heterogeneous volunteers	47.2	22
Homogeneous volunteers	48.7	5
Homogeneous non-volunteers	46.3	30
Total	50.0	444

Table 47. Analysis of Variance Table for What the Adults at Home Think About Student Behavior and General Educational Quality.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	79.05	2	39.53	
Within Tracks	2897.78	441		
Between Groups-Track One	5.02	2	2.51	.38
Between Groups-Track Two	5.76	2	2.88	.43
Between Groups-Track Three	1.85	2	.93	.14
Within Groups Error (Residual)	2885.15	435	6.63**	
Total	2976.83	443	6.72	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{6.63}{6.72}(100) = 98.66$.

Table 48. Items That Define What the Adults at Home Think About School and Ability Grouping.

Item Number	Question	Factor Coefficient
26	The adults at home think there should be tracking in Arithmetic and English. FALSE TRUE	.75
29	The adults at home think there should be special classes for slow learners. FALSE TRUE	.71
18	The adults at home think there should be special classes for fast learners. FALSE TRUE	.68
19	The adults at home think there should be no grouping in social studies. TRUE FALSE	.67
27	The adults at home think all students in a class should be of the same ability. FALSE TRUE	.62

Proportion of variance accounted for by Factor II: 28 percent.

Table 49. Mean Scores of What the Adults at Home Think About Ability Grouping.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	50.5	184
Heterogeneous volunteers	44.6	35
Homogeneous volunteers	47.7	28
Homogeneous non-volunteers	52.9	121
Track Two Students	49.7	203
Heterogeneous volunteers	51.3	29
Homogeneous volunteers	47.4	33
Homogeneous non-volunteers	50.0	141
Track Three Stu-ents	50.0	57
Heterogeneous volunteers	49.4	22
Homogeneous volunteers	44.3	5
Homogeneous non-volunteers	51.4	30
Total	50.0	444

Table 50. Analysis of Variance Table of What the Adults at Home Think About Ability Grouping.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	4.14	2	2.07	
Within Tracks	2863.09	441		
Between Groups-Track One	140.16	2	70.08	11.32*
Between Groups-Track Two	16.91	2	8.46	1.37
Between Groups-Track Three	15.08	2	7.54	1.22
Within Groups Error (Residual)	2690.94	435	6.19**	
Total	2867.23	443	6.47	

*Significant at $\alpha = .05$ $F_{2,435}(.95) = 3.00$

**Mean square within on the mean 50 and standard deviation 10 scale is given by $MSW = \frac{6.19}{6.47}(100) = 95.67$.

Table 51. Items that Define What the Adults at Home Think About School and Its Administration and Faculty.

Item Number	Question	Factor Coefficient
25	The adults at home think the principal at this school is well liked. TRUE FALSE	.72
13	The adults at home think the principal at this school is a strong leader. TRUE FALSE	.71
22	The adults at home think the school tells them about what is really going on. TRUE FALSE	.59
17	The adults at home think complaints about this school are taken seriously. TRUE FALSE	.51
14	The adults at home think my teachers know their subjects. TRUE FALSE	.50

Proportion of variance accounted for by Factor III: 36 percent.

Table 52. Mean Scores of What the Adults at Home Think About School Administration and Faculty.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	50.7	184
Heterogeneous volunteers	49.8	35
Homogeneous volunteers	50.9	28
Homogeneous non-volunteers	51.0	121
Track Two Students	49.0	203
Heterogeneous volunteers	48.6	29
Homogeneous volunteers	48.5	33
Homogeneous non-volunteers	49.2	141
Track Three Students	50.4	57
Heterogeneous volunteers	50.6	22
Homogeneous volunteers	52.5	5
Homogeneous non-volunteers	48.8	30
Total	50.0	444

Table 53. Analysis of Variance Table of What the Adults at Home Think About School Administration and Faculty.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	19.84	2	9.92	
Within Tracks	2392.55	441		
Between Groups-Track One	2.73	2	1.37	.25
Between Groups-Track Two	1.32	2	.66	.12
Between Groups-Track Three	2.18	2	1.09	.20
Within Groups Error (Residual)	2386.32	435	5.49**	
Total	2412.39	443	5.45	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{5.49}{5.45}(100) = 100.73$

stimulating the Track One students in the experimental classes to do well relative to their own ability rather than to do well relative to the class average. The teachers felt that they did not have enough time to prepare adequately before the advent of the first year's study to stimulate all the higher ability students in the experimental classes, but in any case, it should be emphasized that the differences are not significant. The higher scores of the Track Three students in the experimental classes on the Constitution post test is some indication that Track Three students can improve in a more heterogeneous group, at least when the questions asked are of the more concrete textbook, factual nature.

The main conclusions that can be drawn from the first year's study are:

1. Heterogeneous trained students in Tracks Two and Three performed as well as their counterparts in the homogeneous classes on the relevant items of the Cooperative Social Studies Test. Track One students performed slightly below expectation.

2. Students trained in heterogeneous classes did as well on incidental learning for the Cooperative Social Studies Test as did their counterparts trained in homogeneous classes.

3. Heterogeneous volunteers in Track Three showed higher scores on the United States Constitution post test.

4. Heterogeneous volunteers in Track Three showed a self-image that was more critical and more in line with the students in Track One.

5. Heterogeneous volunteers in Track One found their peers as more masculine in their behavior.

6. Heterogeneous volunteers in Track Three found their peers as more masculine in their behavior.

7. Students in the heterogeneous classes were less satisfied with their class assignments during the experimental year.

8. It is quite clear that students who were in the heterogeneous volunteer classes or who did volunteer but were assigned to the homogeneous classes reported that their parents did not look with favor upon ability groupings and tracking.

The Experimental Groups for the 1968-1969 School Year

During the second year of the study, students transferred to the freshman campus of the senior high school. In addition, they were joined by students from the remaining junior high school of the district. These new students were given an opportunity to join the study if their parents so desired. In addition, the students in the study of the previous year were given the opportunity to continue in the new classes. The letters sent to parents are included in this report as Appendix B. (Letters sent the previous year were very similar.) As suggested in this letter, an attempt was made to prepare the teachers for their heterogeneous classes. Much of the preparation was based on what was learned by the teachers in the 1967-1968 study.

From the volunteers, heterogeneous classes were created by selecting students at random. The remaining students were assigned to homogeneous classes.

The remainder of this narrative relates to the attitudes and achievement scores of the second year of the study.

The format of the data presentation and analysis follows that of the first year, except that the analysis becomes more involved because of the inclusion of the new students from Willard Junior High School.

Initial Differences in School Achievement and Ability of the Experimental Groups of the 1968-1969 School Year

At the beginning of the second year's study, all students in the school were given the Differential Aptitude Test. The analysis of variance for the verbal score on this test is shown in Table 55. The corresponding mean standardized scores are shown in Table 54. Except for the differences between the tracks, none of the F-values are significant, suggesting that differences between comparable groups are not significant. Since these tables are more complex and detailed than those of the previous year, these two tables will be examined in greater detail so as to simplify the discussion for subsequent tables.

Track One students averaged 58.2, or almost one standard deviation above the mean score for the 904 students tested. The Track Two students averaged 46.4, while the Track Three students averaged 40.2. These large differences in verbal aptitude between the tracks is not unexpected. Within the Track One students, the heterogeneous volunteers scored 57.0, the homogeneous volunteers scored 58.8, and the homogeneous non-volunteers scored 58.2. As expected, these three means are not statistically significant from one another. For the heterogeneous volunteers, the continuing students scored 55.3 while the new heterogeneous volunteers scored 58.5. In this case, this

Table 54. Mean Scores on the Verbal Items of the Differential Aptitude Test.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	58.2	372
Heterogeneous volunteers	57.0	33
Continuing students	55.3	16
New students	58.5	17
Homogeneous volunteers	58.8	39
Continuing students	57.7	9
New students	59.1	30
Homogeneous non-volunteers	58.2	300
Track Two Students	46.4	350
Heterogeneous volunteers	48.6	33
Continuing students	49.8	16
New students	47.4	17
Homogeneous volunteers	46.1	32
Continuing students	45.6	8
New students	46.2	24
Homogeneous non-volunteers	46.2	285
Track Three Students	40.2	182
Heterogeneous volunteers	41.7	17
Continuing students	42.5	8
New students	41.0	9
Homogeneous volunteers	40.8	10
Continuing students	41.9	6
New students	39.1	4
Homogeneous non-volunteers	40.1	155
Total	50.0	904

Table 55. Analysis of Variance Table for the Verbal Scores on the Differential Aptitude Test.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio
Between Tracks	63101.0	2	31550.5	
Within Tracks	59501.3	901		
Between Groups-Track One	86.6	2	43.3	1.74
Cont. x New-Hetero. Classes	115.5	1	115.5	
Cont. x New-Homo Classes	17.9	1	17.9	
Between Groups-Track Two	230.2	2	115.1	1.73
Cont. x New-Hetero Classes	62.8	1	62.8	
Cont. x New-Homo Classes	3.0	1	3.0	
Between Groups-Track Three	63.2	2	31.6	.48
Cont. x New-Hetero Classes	14.2	1	14.2	
Cont. x New-Homo Classes	26.7	1	26.7	
Within Groups Error (Residual)	58881.2	889	66.23**	
Total	122602.3	903	135.80	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{66.23}{135.80}(100) = 48.77$.

this difference is not significant. Thus, with respect to verbal attitude as measured by this test, there are no differences between the continuing heterogeneous volunteers and the new heterogeneous volunteers coming from the other junior high school in the school district.

As can be seen, none of the corresponding differences in mean scores are statistically significant. Thus, it can be concluded that verbal ability between the students in the three tracks is statistically significant, but within the tracks there are no statistically significant differences between the students placed in the heterogeneous or homogeneous classes. Also, the new students coming into the study have verbal ability scores that are essentially the same as those of the corresponding students in the previous year's study.

In Table 56 the mean scores on the numeric part of the Differential Aptitude Test are shown and in Table 57 is shown the analysis of variance table for the various study groups of the investigation. As can be seen, none of the comparisons are statistically significant except for the new and continuing students in the Track Three classes assigned to the heterogeneous classrooms. The continuing volunteers scored 45.8 on the numeric items while the new volunteers scored 38.4. This 7.4 standard units in score differences is statistically significant. Why this difference should be significant is not clear since the students were assigned at random to the heterogeneous and homogeneous classes. For the Track Three new volunteer students the mean score on the numeric items is given by $\bar{X}_N = \frac{9(38.4) + 4(42.3)}{13} = 39.6$, indicating that the

Table 56. Mean Scores on the Differential Aptitude Test--Numeric Items.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	57.4	372
Heterogeneous Volunteers	57.1	33
Continuing Students	57.0	16
New Students	57.2	17
Homogeneous Volunteers	57.9	39
Continuing Students	56.1	9
New Students	58.6	30
Homogeneous Non-volunteers	57.2	300
Track Two Students	47.0	350
Heterogeneous volunteers	48.7	33
Continuing Students	50.0	16
New Students	47.6	17
Homogeneous Volunteers	46.9	32
Continuing Students	45.8	8
New Students	47.4	24
Homogeneous Non-volunteers	46.8	285
Track Three Students	40.6	182
Heterogeneous Volunteers	41.9	17
Continuing Students	45.8	8
New Students	38.4	9
Homogeneous Volunteers	42.0	10
Continuing Students	41.6	6
New Students	42.3	4
Homogeneous Non-volunteers	40.4	155
Total	5C.0	904

Table 57. Analysis of Variance Table for Differential Aptitude Test: Numeric Items.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	31149.4	2	15574.7	
Within Tracks	40440.3	901		
Between Groups-Track One	11.5	2	5.8	
Cont. x New-Hetero Classes	.3	1	.3	
Cont. x New-Homo Classes	30.5	1	30.5	
Between Groups-Track Two	81.5	2	40.7	
Cont. x New-Hetero Classes	37.9	1	37.9	
Cont. x New-Homo Classes	12.1	1	12.1	
Between Groups-Track Three	39.4	2	19.7	
Cont. x New-Hetero Classes	180.5	1	180.5	4.01*
Cont. x New-Homo Classes	2.4	1	2.4	
Within Groups Error (Residual)	40044.2	889	45.0**	
Total	71589.7	903	79.3	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{45.0}{79.3}(100) = 56.75$.

***F-values less than 1 are not shown.

*Significant at $\alpha = .05$ $F_{1,889}(.95) = 3.84$

overall group is somewhat below the continuing volunteers whose mean score is given by $\bar{X}_C = \frac{8(45.8) + 6(41.6)}{14} = 44.0$.

Mean standardized scores for the spatial relations items of the Differential Aptitude Test for the various experimental groups are shown in Table 58. The corresponding analysis of variance table is shown in Table 59. In a somewhat unexpected manner, the differences between the three tracks are quite large. Track One students average 56.5, Track Two students average 48.3, while Track Three students average 39.6. For Track One and Track Two students none of the internal comparisons are significant. However, the mean difference of $\hat{\Delta} = 40.1 - 37.2 = 2.9$ between the new and continuing students placed in the heterogeneous classes of Track Three is significant. This agrees with the findings for the numberic items on the Differential Aptitude Test.

Summary Statements for the Differential Aptitude Test

Of the 27 F-tests performed on the verbal, the numeric, and spatial relations portions of the Differential Aptitude Test, two comparisons were found to be statistically significant at the .05 level of significance. This is about as close as one could get to a chance number of significant findings of $T = 27(.05) = 1.35$, suggesting that the two significant findings should be ignored. For this reason, it is concluded that the students in the various conditions of the study are well matched with respect to verbal ability, numerical ability, and spatial relations as measured by the Differential Aptitude Test.

Performance in Social Studies Following the Second Year's Study

In mid-May of 1969, all students in the school were

Table 58. Mean Scores on the Differential Aptitude Test--Spatial Relations Items.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	56.5	360
Heterogeneous Volunteers	56.4	31
Continuing Students	56.7	15
New Students	57.2	16
Homogeneous Volunteers	57.6	39
Continuing Students	57.7	9
New Students	57.6	30
Homogeneous Non-volunteers	56.4	290
Track Two Students	48.3	332
Heterogeneous Volunteers	49.1	32
Continuing Students	49.6	16
New Students	48.5	16
Homogeneous Volunteers	49.1	32
Continuing Students	50.6	8
New Students	48.4	24
Homogeneous Non-volunteers	48.1	268
Track Three Students	39.6	169
Heterogeneous Volunteers	41.5	15
Continuing Students	40.1	7
New Students	37.2	8
Homogeneous Volunteers	39.1	10
Continuing Students	42.7	6
New Students	33.6	4
Homogeneous Non-volunteers	39.4	144
Total	50.0	861

Table 59. Analysis of Variance Table for the Spatial Relations Scores on the Differential Aptitude Test.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	44293.0	2	22146.5	
Within Tracks	66393.2	858		
Between Groups-Track One	70.6	2	35.3	
Cont. x New-Hetero Classes	22.6	1	22.6	
Cont. x New-Homo Classes	.1	1	.1	
Between Groups-Track Two	60.6	2	30.3	
Cont. x New-Hetero Classes	13.8	1	13.8	
Cont. x New-Homo Classes	36.3	1	36.3	
Between Groups-Track Three	75.0	2	37.5	
Cont. x New-Hetero Classes	382.7	1	382.7	4.95*
Cont. x New-Homo Classes	264.6	1	264.6	3.42
Within Groups Error (Residual)	65466.9	846	77.4**	
Total	110686.2	860	128.7	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{77.4}{128.7}(100) = 60.14$.

***F values less than 1 are not shown.

*Significant at $\alpha = .05$ $F_{1846}(.95) = 3.84$

given a social studies test which was designed by the teachers acting as a collective body. This included teachers who taught students in heterogeneous classes as well as teachers who taught students in the homogeneous classes. The objectives of the test were to tap the knowledge and skills that the teachers thought should have been covered in the classes and acquired by the students to show that they had successfully learned the material of the curriculum. A copy of this test is presented as Appendix C.

The mean scores for the test are presented in Table 60, while the corresponding analysis of variance is presented in Table 61. As expected, the standardized differences between the tracks accounts for the greatest part of the variability. For this test, $\hat{\omega}^2 = \frac{102446.04}{228542.82} = .45$, indicating that about 45 percent of the total variation in the test scores can be attributed to the differences that exist between the tracks. For the three tracks the mean scores are 56.4, 48.1, and 37.2.

For the Track One students, the F-ratio for the comparison between the heterogeneous volunteers, the homogeneous volunteers, and the homogeneous non-volunteers is given by $F = 1.06$ which indicates that the mean differences are not significant. The corresponding mean scores are 57.9, 56.9, and 56.0. In addition, the differences between the new and continuing students are not significant. For the heterogeneous classes, $F = 1.58$, while for the homogeneous classes, $F < 1$. In this case, the means are 56.2, 59.4, 57.6, and 56.6. As these statistics indicate, performance on the teacher made test is unrelated to the kind of classroom, heterogeneous or

Table 60. Mean Scores for the Teacher-made Test of the Second Year's Study.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	56.4	341
Heterogeneous Volunteers	57.9	33
Continuing Students	56.2	15
New Students	59.4	18
Homogeneous Volunteers	56.9	35
Continuing Students	57.	9
New Students	56.6	26
Homogeneous Non-volunteers	56.0	273
Track Two Students	48.1	316
Heterogeneous Volunteers	52.8	32
Continuing Students	52.6	18
New Students	53.1	14
Homogeneous Volunteers	47.3	31
Continuing Students	44.2	7
New Students	48.2	24
Homogeneous Non-volunteers	47.6	253
Track Three Students	37.2	121
Heterogeneous Volunteers	42.3	17
Continuing Students	43.4	10
New Students	41.4	7
Homogeneous Volunteers	38.2	7
Continuing Students	40.1	4
New Students	35.7	3
Homogeneous Non-volunteers	36.1	97
Total	50.0	778

Table 61. Analysis of Variance Table for the Teacher-made Test of the Second Year.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	102446.0	2	51223.0	
Within Tracks	126096.8	775		
Between Groups-Track One	334.6	2	167.3	1.06
Cont. x New-Hetero Classes	250.5	1	250.5	1.58
Cont. x New-Homo Classes	20.0	1	20.0	
Between Groups-Track Two	2354.3	2	1177.1	7.43*
Cont. x New-Hetero Classes	8.0	1	8.0	
Cont. x New-Homo Classes	256.2	1	256.2	1.62
Between Groups-Track Three	1888.7	2	944.4	5.96*
Cont. x New-Hetero Classes	61.3	1	61.3	
Cont. x New-Homo Classes	94.3	1	94.3	
Within Groups Error (Residual)	120829.0	763	158.4**	
Total	228542.8	777	294.13	

*Significant at $\alpha = .05$ $F_{2,763}(.95) = 3.00$

**Mean square within on the mean 50 and standard deviation 10 scale is given by $MSW = \frac{158.4}{294.13}(100) = 53.85$.

***F values less than 1 are not shown.

homogeneous, that a Track One student attends. This is somewhat different from the results of the first year's study in which it was noted that the Track One students in heterogeneous classes did not do as well as Track One students in homogeneous classes even though the difference was not significant. As was noted, the teachers in the first year reported that motivating the Track One students was somewhat of a problem. This problem seems not to have been that important for the second year study. Since the teachers of the second year study knew of this problem, they made extra efforts to motivate these children. On this basis, it is reasonable to conclude that performance of Track One students is unrelated to the ability make-up of the classroom, provided that teachers take the effort to motivate Track One students to learn and act in a fashion commensurate with their favorable intellectual endowment. Thus, Hypothesis 1 is supported for this particular achievement measure.

For Track Two students the findings are quite different. In this case, the F-ratio for the comparisons of heterogeneous volunteers, homogeneous volunteers, and homogeneous non-volunteers is statistically significant since $F = 7.43$. The mean scores for these three groups are 52.8, 47.3, and 47.6, respectively. For the heterogeneous trained Track Two students versus the homogeneous trained Track Two students, the mean difference is given by $\hat{\Delta} = 52.8 - \frac{31(47.3) + 253(47.6)}{284} = 52.8 - 47.6 = 5.2$, which is more than one-half standard deviation. Within the heterogeneous and homogeneous classes the differences between the continuing students and the new students

are not significant. The F-ratios for the heterogeneous and homogeneous classes are given by $F < 1$ and $F = 1.62$. Thus, it can be concluded that for Track Two students, training in a heterogeneous or homogeneous class does make a difference. In terms of standardized units, students trained in a heterogeneous class outscore their Track Two counterparts trained in homogeneous classes by .5 of a standard deviation. Thus, Hypothesis 2 of the study is supported for this achievement measure.

For Track Three students the differences involving the heterogeneous volunteers, the homogeneous volunteers, and the homogeneous non-volunteers are significant. The corresponding mean scores are given by 42.3, 38.2, and 36.1 with the F-ratio given by $F = 5.96$. For the heterogeneous trained students versus the homogeneous trained students, the mean difference is given by $\hat{\Delta} = 42.3 - \frac{7(38.2) + 97(36.1)}{104} = 42.3 - 36.2 =$ 6.1 standardized units. Thus, it is concluded that the Track Three students trained in heterogeneous classes outperform the Track Three students trained in homogeneous classes by more than .6 of a standard deviation and that Hypothesis 3 is supported. For educational variables, differences of this magnitude are indicative of a major statistical finding.

For the Cooperative Social Studies Test, the results are not so impressive even though the differences are in the same direction as those noted for the teacher-made test. In any case, none of the differences are statistically significant as is noted by examination of Tables 62 and 63 which report the mean performances for the various groups and the analysis of

Table 62. Mean Scores on the Cooperative Social Studies Test.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	55.0	238
Heterogeneous Volunteers	57.2	26
Continuing Students	54.9	12
New Students	57.1	14
Homogeneous Volunteers	54.3	26
Continuing Students	57.4	6
New Students	53.4	20
Homogeneous Non-volunteers	56.5	186
Track Two Students	47.9	195
Heterogeneous Volunteers	51.8	21
Continuing Students	52.5	9
New Students	51.2	12
Homogeneous Volunteers	47.0	16
Continuing Students	41.1	4
New Students	49.1	12
Homogeneous Non-volunteers	47.6	158
Track Three Students	36.3	66
Heterogeneous Volunteers	38.1	14
Continuing Students	36.5	7
New Students	39.7	7
Homogeneous Volunteers	33.6	5
Continuing Students	34.8	2
New Students	32.8	3
Homogeneous Non-volunteers	36.1	47
Total	50.0	499

Table 63. Analysis of Variance Table for the Cooperative Social Studies Test.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	37733.8	2	18866.9	
Within Tracks	54279.5	496		
Between Groups-Track One	73.6	2	36.8	
Cont. x New-Hetero Classes	109.6	1	109.6	1.00
Cont. x New-Homo Classes	132.9	1	132.9	1.21
Between Groups-Track Two	626.2	2	313.1	2.87
Cont. x New-Hetero Classes	17.3	1	17.3	
Cont. x New-Homo Classes	341.3	1	341.3	3.13
Between Groups-Track Three	151.5	2	75.7	
Cont. x New-Hetero Classes	64.3	1	64.3	
Cont. x New-Homo Classes	8.5	1	8.5	
Within Groups Error (Residual)	52754.3	484	109.0**	
Total	92013.3	498	184.77	

** Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{109.00}{184.77}(100) = 58.99$.

*** F values less than 1 are not shown.

variance for the scores of the test. Even though the results are not statistically significant at the .05 level, exactly the same findings as noted for the teacher-made test are significant at the .10 level. If the sample sizes for the Cooperative Social Studies Test had been the same as those for the teacher-made test, the results would probably have been statistically significant at the .05 level.

As with the teacher-made test, most of the variability between the various experimental groups can be associated with the differences between the various tracks. The amount of explained variance for this variable is given by $\hat{\omega}^2 = \frac{37733.8}{92013.3} = .41$ so that the variability between the tracks accounts for 41 percent of the total variability. The mean standardized scores for the three tracks are 55.0, 47.9, and 36.3. These are quite close to the standardized scores on the teacher-made test of 56.4, 48.1, and 37.2. For the three experimental groups of Track Two, the F value is given by F = 2.87 which is significant at the .10 level. The mean scores for the heterogeneous volunteers, the homogeneous volunteers, and the homogeneous non-volunteers are equal to 51.8, 47.0, and 47.6, indicating that the students placed in the heterogeneous classes out-perform their homogeneous counterparts by about four standardized units.

Finally, at the 10 percent level of significance it would be concluded that the new volunteers of Track Two placed in homogeneous classes outperformed the continuing students by 8.0 standardized points. Since only four students were in the continuing groups, this finding should be weighed with some

caution.

For the Track Three students, it is seen that the heterogeneous trained students did better than the homogeneous trained students. The mean difference is given by $\hat{\Delta} = 38.1 - \frac{5(33.6) + 47(36.1)}{52} = 38.1 - 35.8 = 2.3$, but because the sample sizes are small, the chances of identifying this difference as statistically significant is so small as to be negligible.

Thus, while none of the major differences are significant at the .05 level of significance, all are in the direction noted for the teacher-made test and would cause one to think that the findings for this variable repeat what one would find if a larger sample had been available for analysis.

Attitudes Towards Classwork, Self, and Other Students in Social Studies During the 1968-1969 School Year

During the second year of the study, students were given an attitude inventory at the beginning of the school year. This inventory was considerably shorter than the one given during the first year mainly because the first year's questionnaire was revised and improved by discarding nondiscriminating items. Since the results of the pre test parallel the post test of the first year, the results are not presented and discussed. Instead, the post test results for the second year are examined in some detail. In making this evaluation, it should be recalled that a large number of students had completed two full years in heterogeneous classes. If being placed in a heterogeneous class induces fundamental changes in attitude, then there was sufficient time for such changes to have

occurred. As will be noted, few changes did occur suggesting that ability mixing has no strong impact on student's attitudes. For interested readers, the revised attitude inventory is presented in Appendix D.

The adjective pairs that define attitude towards self are reported in Table 64. As before, the first factor, generated from the factor analysis is a factor in self evaluation. The mean scores of the various study groups of the study are shown in Table 65, while the corresponding analysis of variance table is shown in Table 66. As expected, the evaluation of self varies across the three tracks; F = 5.40 represents a significant difference at the .05 level. The mean standardized scores for the three tracks are 48.7, 50.7, and 52.0. Thus, similar to the first year study, the Track One students are more critical of themselves than are the Track Three students. However, the degree of self criticism is not as variable as the previous year where the mean scores were given by 48.3, 50.9, and 54.2.

The movement to a more uniform self evaluation across tracks is somewhat surprising, especially when it is noted that most of the classes in other academic subjects in the second year study were homogeneous. Among the Track Two students, the differences in attitude are statistically significant. The homogeneous volunteers scored 44.6, while the remaining students in the other two groups scored 51.2 and 51.3. Most of this poor scoring can be directly attributed to the 17 students who constitute the new volunteers. Their mean score was only 42.1.

Table 64. Adjective Pairs that Define Factor I: How I Feel About Myself.

Adjective Pair			Factor Coefficient
nice	-	awful	.787
pleasant	-	unpleasant	.744
valuable	-	worthless	.736
good	-	bad	.724
hopeful	-	hopeless	.694
happy	-	sad	.664
top	-	bottom	.657
smart	-	dumb	.587
honest	-	dishonest	.580
active	-	passive	.550

Proportion of variance accounted for by Factor I: 23.1 percent

Table 65. Mean Scores on Factor I: Attitude Towards Self.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	48.7	308
Heterogeneous Volunteers	50.1	29
Continuing Students	52.3	15
New Students	47.7	14
Homogeneous Volunteers	46.6	36
Continuing Students	50.5	9
New Students	45.3	27
Homogeneous Non-volunteers	48.9	243
Track Two Students	50.7	255
Heterogeneous Volunteers	51.2	29
Continuing Students	50.6	13
New Students	51.7	16
Homogeneous Volunteers	44.6	22
Continuing Students	53.3	5
New Students	42.1	17
Homogeneous Non-volunteers	51.3	204
Track Three Students	52.0	106
Heterogeneous Volunteers	48.9	17
Continuing Students	51.7	9
New Students	51.6	8
Homogeneous Volunteers	55.9	9
Continuing Students	53.7	5
New Students	59.1	4
Homogeneous Non-volunteers	48.7	80
Total	50.0	669

Table 66. Analysis of Variance Table for Attitude Towards Self.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	1048.6	2	524.3	
Within Tracks	62745.3	666		
Between Groups-Track One	221.0	2	110.5	1.13
Cont. x New-Hetero Classes	158.1	1	158.1	1.62
Cont. x New-Homo Classes	182.4	1	182.4	1.88
Between Groups-Track Two	886.1	2	443.0	4.56*
Cont. x New-Hetero Classes	8.7	1	8.7	
Cont. x New-Homo Classes	480.5	1	480.5	4.95*
Between Groups-Track Three	424.9	2	212.4	2.18
Cont. x New-Hetero Classes	.1	1	.1	
Cont. x New-Homo Classes	64.3	1	64.3	
Within Groups Error (Residual)	63429.2	654	97.0**	
Total	63793.9	668	95.49	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{97.00}{95.49}(100) = 101.58$.

***F values less than 1 are not shown.

* Significant at $\alpha = .05$ $F_{2,654}(.95) = 3.00$

$F_{1,654}(.95) = 3.84$

Similar to the finding for the Track One students, none of the differences between the groups for the Track Three students are significant.

The adjective pairs that define attitudes towards other kids in this school are shown in Table 67. Unlike the previous year's study in which this factor was more an evaluation of behavior, this factor is very much like the factor generated for the evaluation of self. The major difference is in the presence of the adjective pair (clean-dirty) which was also present in the previous year's factor that defined how the students felt about the other kids.

The mean scores for this factor are shown in Table 68, while the analysis of variance table is shown in Table 69. For the differences between the tracks, the F-value is given by F = 5.72 with the mean scores for the three tracks given by 51.6, 49.0, and 49.4. While the differences are significant, they are so small as to be almost meaningless. Except for the new and continuing volunteers placed in the heterogeneous classes, none of the other differences are statistically significant. Why this one difference is significant is not clear. It could represent a Type I error.

The factor analysis for the 20 adjective pairs gives rise to another factor characterized by the adjective pairs shown in Table 70. These pairs seem to describe a good or bad guy characteristic of the students. A bad guy or troublemaker is seen as bully, unfair, bitter, dishonest, distasteful, and ugly, while a good guy, or goodie guy, is seen as gentle, fair, sweet, honest, tasteful, and beautiful. The mean scores for

Table 67. Adjective Pairs That Define Factor II: How I Feel About the Others.

Adjective Pair			Factor Coefficient
nice	-	awful	.822
valuable	-	worthless	.804
pleasant	-	unpleasant	.803
good	-	bad	.801
hopeful	-	hopeless	.800
top	-	bottom	.753
smart	-	dumb	.733
clean	-	dirty	.673
happy	-	sad	.523

Proportion of variance accounted for by rotated factor: 26.6 percent.

Table 68. Mean Scores for: How I Feel Towards Others.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	51.6	308
Heterogeneous Volunteers	53.0	29
Continuing Students	52.0	15
New Students	54.1	14
Homogeneous Volunteers	54.3	36
Continuing Students	54.5	9
New Students	54.2	27
Homogeneous Non-volunteers	51.1	243
Track Two Students	49.0	255
Heterogeneous Volunteers	49.5	29
Continuing Students	49.1	13
New Students	49.8	16
Homogeneous Volunteers	49.6	22
Continuing Students	53.8	5
New Students	41.3	17
Homogeneous Non-volunteers	48.9	204
Track Three Students	49.4	106
Heterogeneous Volunteers	48.9	17
Continuing Students	44.0	9
New Students	54.4	8
Homogeneous Volunteers	55.9	9
Continuing Students	56.4	5
New Students	55.1	4
Homogeneous Non-volunteers	48.7	80
Total	50.0	669

Table 69. Analysis of Variance Table for How I Feel Towards Others.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	1056.7	2	528.4	
Within Tracks	63191.1	666		
Between Groups-Track One	380.6	2	190.3	2.06
Cont. x New-Hetero Classes	30.4	1	30.4	
Cont. x New-Homo Classes	.8	1	.8	
Between Groups-Track Two	18.6	2	9.3	
Cont. x New-Hetero Classes	3.3	1	3.3	
Cont. x New-Homo Classes	213.1	1	213.1	2.30
Between Groups-Track Three	424.8	2	212.4	2.30
Cont. x New-Hetero Classes	460.3	1	460.3	4.98*
Cont. x New-Homo Classes	4.8	1	4.8	
Within Groups Error (Residual)	61654.4	654	92.3**	
Total	64247.8	668	96.18	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{92.3}{96.18}(100) = 95.97$.

***F values less than 1 are not shown.

*Significant at $\alpha = .05$ $F_{1,654}(.95) = 3.84$

Table 70. Adjective Pairs That Define Factor III: How I Feel About the Other Kids in This School.

Adjective Pair			Factor Coefficient
gentle	-	bully	.813
fair	-	unfair	.740
sweet	-	bitter	.751
honest	-	dishonest	.729
tasteful	-	distasteful	.699
beautiful	-	ugly	.658

Proportion of variance accounted for by rotated factor: 18.5 percent.

the various groups on this factor are shown in Table 71, while the corresponding analysis of variance table is shown in Table 72. According to the F-values reported in this latter table, none of the comparisons are statistically significant, indicating that attitude on this dimension is unrelated to track assignment and the training in heterogeneous or homogeneou. classes.

The items that define attitude towards classwork and assignments in social studies during the second year's investigation are shown in Table 73. It is interesting to note that these same items clustered together for the analysis presented in Table 33. The mean scores for the various groups are presented in Table 74 and the corresponding analysis of variance is presented in Table 75. Outside of the differences between tracks, none of the comparisons involving heterogeneous versus homogeneous classes are significant. Thus, assignement in heterogeneous or homogeneous classes has no impact upon the way students feel about classwork or assignments during the second year of the study with respect to social studies. As expected, the students in Track Three show the most dissatisfaction with the assignments and classwork in the social studies class; however, the dissatisfaction is not very strong.

Attitudes of the Adults at Home to the School Environment as Perceived by the Student

Concerning what the parents at home think, two factors were generated by the factor analysis. The items that define these factors are shown in Tables 76 and 79. The first factor refers to the educational quality of the school while the second

Table 71. Mean Scores for Factor III: How I Feel About the Other Kids in this School.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	51.1	311
Heterogeneous Volunteers	51.9	29
Continuing Students	52.9	15
New Students	50.9	14
Homogeneous Volunteers	54.4	36
Continuing Students	53.8	9
New Students	54.6	27
Homogeneous Non-volunteers	50.4	246
Track Two Students	49.2	258
Heterogeneous Volunteers	50.4	29
Continuing Students	49.6	13
New Students	51.1	16
Homogeneous Volunteers	49.1	22
Continuing Students	44.8	5
New Students	50.4	17
Homogeneous Non-volunteers	49.0	207
Track Three Students	49.5	104
Heterogeneous Volunteers	47.6	17
Continuing Students	44.2	9
New Students	51.4	8
Homogeneous Volunteers	56.6	8
Continuing Students	56.1	5
New Students	57.6	3
Homogeneous Non-volunteers	49.2	79
Total	50.0	673

Table 72. Analysis of Variance Table for Factor III: How I Feel About the Other Kids in this School.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	543.7	2	271.9	
Within Tracks	61175.4	670		
Between Groups-Track One	501.2	2	250.6	2.75
Cont. x New-Hetero Classes	29.5	1	29.5	
Cont. x New-Homo Classes	4.5	1	4.5	
Between Groups-Track Two	47.2	2	23.6	
Cont. x New-Hetero Classes	16.5	1	16.5	
Cont. x New-Homo Classes	118.3	1	118.3	1.30
Between Groups-Track Three	475.0	2	237.5	2.61
Cont. x New-Hetero Classes	223.9	1	223.9	2.47
Cont. x New-Homo Classes	4.3	1	4.3	
Within Groups Error (Residual)	59755.0	658	90.8**	
Total	61719.1	672	91.84	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{90.8}{91.84}(100) = 98.87$.

***F values less than 1 are not shown.

Table 73. Items That Define Attitudes Towards Classwork and Assignments in Social Studies for the Second Year.

Item No.	Statement	Factor Coefficient
3	Do not like to come to class	.796
9	Not happy to do assignments	.762
1	Do not like classwork	.760
6	Do not like teacher	.720
5	Do not learn	.678
7	Do not enjoy extra work	.624

Proportion of variance accounted for by rotated factor: 24.3 percent.

Table 74. Mean Scores for Attitudes Towards Assignments and Classwork in Social Studies This Year.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	48.1	322
Heterogeneous Volunteers	48.0	30
Continuing Students	47.2	16
New Students	48.8	14
Homogeneous Volunteers	47.5	37
Continuing Students	48.1	10
New Students	47.5	27
Homogeneous Non-volunteers	48.2	255
Track Two Students	50.4	259
Heterogeneous Volunteers	48.0	30
Continuing Students	49.3	14
New Students	47.0	16
Homogeneous Volunteers	51.3	23
Continuing Students	50.7	5
New Students	51.4	18
Homogeneous Non-volunteers	50.6	206
Track Three Students	52.9	101
Heterogeneous Volunteers	56.0	17
Continuing Students	57.8	9
New Students	54.0	8
Homogeneous Volunteers	54.6	5
Continuing Students	55.7	2
New Students	53.9	3
Homogeneous Non-volunteers	52.1	79
Total	50.0	682

Table 75. Analysis of Variance Table for Attitudes Toward Assignments and Classwork for the Second Year.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	1909.0	2	954.5	
Within Tracks	66094.0	679		
Between Groups-Track One	12.3	2	6.2	
Cont. x New-Hetero Classes	19.8	1	19.8	
Cont. x New-Homo Classes	2.8	1	2.8	
Between Groups-Track Two	193.6	2	96.8	
Cont. x New-Hetero Classes	38.4	1	38.4	
Cont. x New-Homo Classes	2.0	1	2.0	
Between Groups-Track Three	232.2	2	116.1	1.18
Cont. x New-Hetero Classes	63.3	1	63.3	
Cont. x New-Homo Classes	3.9	1	3.9	
Within Groups Error (Residual)	65525.7	667	98.2**	
Total	68003.0	681	99.99	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{98.2}{99.99}(100) = 98.19$.

***F values less than 1 are not shown.

factor refers to the use of tracking and ability groupings. The mean scores for the various comparison groups of the study are shown in Tables 77 and 80, while the corresponding analysis of variance tables are shown in Tables 78 and 81. Concerning educational quality, there are no differences between the three tracks of students and within the tracks there are no statistically signficant differences. However, for the attitudes related to ability groupings there are signficant differences between the groups with F = 14.92. According to the reported responses, students of Track One report their parents are most in favor of ability groupings and tracking while parent of students in Track Three classes are reported as being opposed.

Summary Conclusions for the 1968-1969 School Year

1. Students in the various experimental conditions of the study were well matched at the beginning of the school year with respect to verbal ability, numerical ability, and spatial relations as measured by the Differential Aptitude Test. While there were large differences between the students in the three different tracks, the differences between the groups within the tracks were not significant.

2. For the teacher-made test, performance for Track One students was independent of the ability group spread within the classes. Students placed in heterogeneous classes did as well as students placed in homogeneous classes. The mean standardized scores for the heterogeneous volunteers was 57.9, while the mean scores for the homogeneous volunteers and non-volunteers were given by 56.9 and 56.0, respectively.

3. For the teacher-made test, performance of the

Table 76. Items that Define What the Parents at Home Think About the School and Its Educational Quality.

Item No.	Statement	Factor Coefficient
4	School Not below standard	.675
11	I should not go to a different school	.656
10	Teachers are well prepared	.649
6	Teachers treat me fairly	.586

Proportion of variance accounted for by Rotated Factor: 15.6 percent.

Table 77. Mean Scores for What the Parents at Home Think About the School and Its Educational Quality.

Experimental Group of the Study	Mean Score	Sample Size
Track One Students	49.8	261
Heterogeneous Volunteers	50.0	26
Continuing Students	48.4	12
New Students	51.3	14
Homogeneous Volunteers	51.8	26
Continuing Students	52.6	6
New Students	51.6	20
Homogeneous Non-volunteers	49.5	209
Track Two Students	49.7	222
Heterogeneous Volunteers	49.9	21
Continuing Students	48.1	9
New Students	51.2	12
Homogeneous Volunteers	49.3	16
Continuing Students	48.5	4
New Students	49.5	12
Homogeneous Non-volunteers	49.7	185
Track Three Students	50.0	83
Heterogeneous Volunteers	50.9	14
Continuing Students	50.6	7
New Students	51.2	7
Homogeneous Volunteers	51.8	5
Continuing Students	52.5	2
New Students	51.3	3
Homogeneous Non-volunteers	49.7	64
Total	50.0	566

Table 78. Analysis of Variance Table for What the Parents at Home Think About the School and Its Educational Quality.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	6.1	2	3.1	
Within Tracks	51871.8	563		
Between Groups-Track One	124.2	2	62.1	
Cont. x New-Hetero Classes	54.6	1	54.6	
Cont. x New-Homo Classes	4.4	1	4.4	
Between Groups-Track Two	3.9	2	1.9	
Cont. x New-Hetero Classes	50.4	1	50.4	
Cont. x New-Homo Classes	3.2	1	3.2	
Between Groups-Track Three	33.6	2	16.8	
Cont. x New-Hetero Classes	1.4	1	1.4	
Cont. x New-Homo Classes	1.8	1	1.8	
Within Groups Error (Residual)	51598.2	551	93.6**	
Total	51877.9	565	91.82	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{93.6}{91.82}(100) = 101.94$.

***F values less than 1 are not shown.

Table 79. Items That Define What Parents at Home Think About Teaching and Ability Grouping.

Item No.	Statement	Factor Coefficient
3	No special class for smart kids	.737
9	No tracking in Math and English	.734
12	No special class for slow kids	.558
5	No tracking in social studies	.529

Proportion of variance accounted for by rotated factor: 15 percent.

Table 80. Items That Define What Parents at Home Think About Tracking and Ability Grouping.

Experimental Group of The Study	Mean Score	Sample Size
Track One Students	47.4	261
Heterogeneous Volunteers	50.3	26
Continuing Students	50.0	12
New Students	50.6	14
Homogeneous Volunteers	49.9	26
Continuing Students	52.6	6
New Students	49.1	20
Homogeneous Non-volunteers	46.7	209
Track Two Students	51.0	222
Heterogeneous Volunteers	52.5	21
Continuing Students	52.5	9
New Students	52.5	12
Homogeneous Volunteers	50.3	16
Continuing Students	50.2	4
New Students	50.3	12
Homogeneous Non-volunteers	50.9	185
Track Three Students	53.4	83
Heterogeneous Volunteers	54.7	14
Continuing Students	56.3	7
New Students	53.2	7
Homogeneous Volunteers	56.4	5
Continuing Students	61.2	2
New Students	53.2	3
Homogeneous Non-volunteers	52.9	64
Total	50.0	586

Table 81. Analysis of Variance Table for What Parents at Home Think About Teaching and Ability Grouping.

Source of Variation	Sum of Squares	d/f	Mean Sq.	F-ratio***
Between Tracks	2915.6	2	1457.8	
Within Tracks	50711.6	563		
Between Groups-Track One	474.6	2	237.3	2.42
Cont. x New-Hetero Classes	2.1	1	2.1	
Cont. x New-Homo Classes	58.0	1	58.0	
Between Groups-Track Two	56.4	2	28.2	
Cont. x New-Hetero Classes	0.0	1	0.0	
Cont. x New-Homo Classes	.1	1	.1	
Between Groups-Track Three	84.9	2	42.5	
Cont. x New-Hetero Classes	35.2	1	35.2	
Cont. x New-Homo Classes	76.4	1	76.4	
Within Groups Error (Residual)	49923.9	511	97.7**	
Total	53627.2	565	94.92	

**Mean square within on the mean 50 and standard deviation 10 scale is given by MSW = $\frac{97.7}{94.92}(100) = 102.93$.

***F values less than 1 are not shown.

Track Two students was related to the class in which their training occurred. Students placed in heterogeneous classes out-performed their homogeneous counterparts by 5.2 standardized units. The mean score for the heterogeneous trained students was 52.8 while the mean score for the homogeneous trained students was 47.6.

4. For the teacher-made test, performance of the TrackThree students was related to the ability make-up of the class in which instruction took place. The mean score for the students assigned to the heterogeneous classes was 42.3 while the mean score for the students trained in the homogeneous classes was 36.2.

5. For the Cooperative Social Studies Test, none of the major differences were significant at the .05 level, though all are in the direction noted for the teacher-made test. Since the sample sizes for this test were considerably smaller, it is quite possible that if a larger sample had been available, the decisions made with respect to the teacher-made test would also be made with respect to this test.

6. As was noted for the first year study, Track One students are more critical of themselves than are the students of Track Two and Track Three.

7. For the self evaluation measures of the second year study, the variability is less than the first year study.

8. When students are asked to report how they feel about other kids in their social studies classes, it was found that their attitudes were not affected by training in heterogeneous or homogeneous classes.

9. When asked to report how they felt about the class assignments and school work, it was found that these attitudes were unrelated to training in heterogeneous or homogeneous classes.

10. Students in Track One report that their parents favor ability grouping while students in Track Three report that their parents are not in favor of ability grouping.

Discussion

The major findings of this study are that student attitudes are not affected to any pronounced degree by training in a heterogeneous or homogeneous class. Attitudes of an individual towards self, classmates, and school seem to have little relationship to the ability of the remaining students in the class. However, the same conclusions do not apply to individual school achievement. In this study the performance of Track One students was not affected by training in heterogeneous or homogeneous social studies classes, at least as measured by the tests of this investigation. On the other hand, students of Tracks Two and Three performed at a higher level in a class that was heterogeneous with respect to ability. Possible reasons for this finding are numerous. The most frequently reported reasons for this finding are that the low achieving students are exposed to models that can be emulated and that the opportunity to interact with high achieving students increases the richness of the environment by expanding the variety of learning situations and experiences for them. Both of these explanations were discussed by Coleman.

According to Coleman (pp. 302-304), "Attributes of

other students account for far more variation in the achievement of minority group children than do any attributes of school facilities and slightly more than do attributes of staff.

"In general as the educational aspirations and background of fellow students increases, the achievement of minority group children increases... The results suggest ... that the environment provided by the student body is asymmetric in its effects, that it has greatest effect on those from educationally deficient backgrounds." In addition it is reported that (pp. 307-308), "...as the proportion white in a school increases, the achievement of students in each racial group increases."

Since Track Three students tend to be Negro and Track One students tend to be white, the results of this investigation suggest that Coleman's comments about schools have a direct parallel to classrooms. The academic achievement of Track Three and Track Two students was higher in the mixed classrooms, while Track One students were essentially unaffected. Both of these classroom findings are in agreement with Coleman's statements about schools.

Another explanation for these findings might be related to the quality of the teachers and what they did in the classrooms while teaching the students with variable abilities. At the beginning of the study, it was decided to give the teachers free reign over their classes and as such, they were told to conduct their classes in what they thought was the best way to teach their pupils. They were supplied with the regular text books for Tracks One, Two, and Three social studies classes. Students were to be permitted to

choose the level of text they wanted to use. While the curriculums by texts across tracks were essentially identical, their presentations were at different levels of complexity, depth, vocabulary, and interest. In addition to multiple texts, the usual maps, films, globes, related texts, newspapers, magazines, and other educational hardware were available to the teachers for class use if so desired. As this suggests, the experimental procedures used in the heterogeneous as well as those used in the homogeneous classes were unstructured and left to the interests and imaginations of the teachers. This procedure was encouraged and followed because it is believed that this is the way the teaching would normally take place in a typical classroom. This means that teachers were free to decide for themselves the manner in which they wanted to conduct their class. What biassing effect this could have had on the outcomes of this study are unclear. If it accounts for a "true" improvement in student achievement, then mixing the tracks has been effective. However, if it only produced a Hawthorne effect, then the outcomes are not so striking. Only future research can show which is the case.

In a natural school setting, teachers, for the most part, are given a curriculum to teach. The manner in which the material is presented is up to the teacher. This study was performed under this model so that generalizations to a broad set of teaching strategies could be achieved. As expected, some teachers were more successful than others in their ability to teach the curriculum in classes with divergent student ability and interest. For that reason, teacher vari-

ables were not controlled in this study nor were they extracted from the analysis of variance as a source of variance. Thus, the teacher variance is always included in the residual or within error term. Since this source of variance tends to inflate the error term, it follows that any significant F-ratio is underestimated relative to the true F-ratio. Thus, all significant statistical findings are identified under the least favorable circumstances. Since the error term is inflated, it also follows that some statistically significant findings were missed.

Because of the unstructured experimental conditions, the immediate question is what did the teachers do in the experimental classes that was different from what they did in the regular tracked classes? Perhaps the best way to answer this question is to let the teachers report for themselves. Here are statements for two of the teachers involved in the 1967-1968 study.

Teacher One

> In the spring of 1968 I taught an experimental heterogeneous class for one six-week grading period. At that time the subject being studied was the federal government system. I used two texts which covered the same material. One of the texts was the book, Civics in Action which was being used in Tracks One and Two. The other that I also used in Tracks Three and Four was a paperback of the Federal Constitution with commentary. I prepared dittoed assignment sheets with parallel content from both books. Students chose which assignment they wanted to do. (Many did both assignments.)
>
> A project was also assigned to all students. They were allowed to approach a subject of their own choice that related to our study in many different ways: scrapbooks, poetry, written reports, oral reports, tape-recorded interviews, and drawings. All members of the class turned in a project.
>
> A great deal of class discussion was used. The

amount of discussion in the heterogeneous class was about the same as that being carried on in Tracks One and Two. There was a good deal more discussion than was taking place in Tracks Three and Four.

On tests there were two groups of questions. The first group was to be answered by all students. In the second group, students were to choose the questions they wanted to answer. For example, they would be required to answer any three questions out of six. The questions ranged in difficulty from simple factual items to those involving speculation and application of material learned. Extra credit points were given to any student who answered more than the required number of questions.

These were my main methods of teaching as I remember them.

Teacher Two

During the 1967-1968 school year we had a pilot study at M. L. King Jr. High School (then Garfield Jr. High School) for the teaching of heterogeneous classes in the area of the eighth grade history of the United States. One of my classes was in the study. The other four remained tracked.

How did I do things differently with this heterogeneous class?

We had available a small number of books of similar content to the text but at an easier reading level for the use of the students. We also had books (2 sets) in sufficient number for each student to use in the classroom. These were in addition to the regular text book. The problem here was to avoid tracking within the classroom.

We did much work through a wide range of projects so that each child could select a project that was suitable for his interest and ability. Group work was encouraged.

Various methods were used to reinforce classroom reading--maps, charts, study sheets, movies, etc. These could be worked on cooperatively to encourage the development of an attitude of helpfulness among the students.

Parts of the tests were planned to be easy enough for all and they also included materials of increasing difficulty to challenge the more capable. This way, reasonable success was within reach of all the students.

Oral discussion and current events gave all the students opportunity in classroom participation.

As I re-read this, I realize that not all these

> methods are different from those used in the other classes--but the emphasis was different--more on cooperation and helpfulness between the students.

Some of the main points that are brought out in these statements are as follows:

1. Preparation of assignments--Teachers reproduced materials for students, materials designed for two reading levels. The students were allowed to choose either assignment. The teachers indicate that many students did both assignments.

2. Reading materials--Besides the regular texts, supplementary books for students of varying abilities were used.

3. Discussions--There apparently was more class discussion in the pilot classes than in the regular Track Three and Four classes and also there was greater student participation.

4. Projects--Various suggestions for projects were made to accommodate the students' abilities and interests. (Scrapbooks, written reports, etc.)

5. Group work--Students were encouraged to work in groups and to help each other.

6. Tests--These were designed for students of all abilities. Questions were both required and optional, with extra credit given for the latter. More difficult questions were used to challenge the more able students.

As these comments suggest, the teacher's activities in the heterogeneous classes were not much different from their activities in a homogeneous class, but what seems to

differentiate the heterogeneous class from the homogeneous class is the interaction that took place between the students. As suggested by one of the teachers, the emphasis on methods was di ferent, with more cooperation and helpfulness between the students. While this may not have been too effective for improving school learning during the first year of the study, one might suspect it was for the Track Two and Three students during the second year of the study.

As to what the teachers did during the second year of the study in the classrooms, information is not clear and where it is available, it seems that the activities were not uniform. Judging from the comments made by the teachers of the first year study and the improved achievement scores of the students during the second year, one might assume that more ideal teaching circumstances were encountered during the second year of the study. As will be noted, the comments of three of the involved teachers does support this supposition. The following illustrates what these teachers reported:

Teacher Three (Edited comments)

All students used the same text though the texts may have differed on different days according to the class assignment and topic. Study groups worked effectively with the brighter students helping the less able. All students could read at the seventh grade level so that non-readers were of minimal difficulty.

There were no arguments or difficulties over class assignments and discipline problems were minimal. Due dates were met on time. Quiet prevailed within the class and stu-

dents seemed to show respect for one another. They showed more courtesy towards one another than shown by students in the homogeneous classes. Students would help one another on class assignments. There was some resistance to this at the beginning but it was not encountered later.

There was grouping within the class, but it was based on interest and not on ability. Visual aids were used once a week as was done in the homogeneous classes. However, students took more interest in them and they were more motivated to produce their own visual aids.

The course materials were structured according to the curriculum guide developed in the summer. More individual topics were covered in the homogeneous classes than in the heterogeneous classes because there was less confusion in the Track One and Track Two classes. Because the heterogeneous class had a large top and middle, it seemed to move smoothly through the curriculum.

Class assignments were uniform across the students. Some assignments were conducted in groups in the class. Students helped each other and took turns leading class discussions, but no student tutoring occurred. Class projects were not uniform and special credits were given for extra work.

The class remained stable. Students knew they could not check out to another class. The parents were involved in enrollment in the test class. There was a spirit of cooperation among the students. They all tried to make the experience successful. They praised each other when they were proud of a project. They were not unkind to each other. Homogeneous

classes were more peaceful, but less enthusiasm was noted among the students. The homogeneous classes seemed to be less fun for the students. A family feeling came out of the heterogeneous class.

Teacher Four (Edited comments)

Three different text books were used. An attempt was made to control the class atmosphere by increasing the number of class discussion and person to person activities. Some activities were grouped across abilities so that activity groups consisted of fast, medium, and slow students. There was considerably less use of audio-visual materials and greater reliance upon class discussions. The assignments in the heterogeneous classes were more varied than those in the homogeneous classes because of the greater variety of students. More of the projects were non-verbal in nature. All students were given minimal assignments that were done individually. Bright students were used as tutors for slow students; however, this program had limited success. Use was made of a teacher aide who was able to follow-up on slower students or was able to induce brighter students into extra activities and work.

The idea that this was a special class seemed to affect the students positively. One student normally assigned to Track One and quite sensitive to the situations existing in a heterogeneous class said, "If this is how being a guinea pig is, I'd like to be one every year."

It is my belief that one must work on atmosphere and human relationships. There is a need for teachers to re-evaluate their techniques and reasons for teaching certain

things. There is greater need for planning time and in-service training during regular school hours and not on evenings and weekends.

Teacher Five (Edited comments)

The following is a summary of some of my findings and feelings concerning the pilot heterogeneous history class of the 1968-1969 school year:

Class Structure: There naturally was a greater range of ability in the pilot heterogeneous class than in the homogeneous classes, but I had the feeling that the pilot class was not truly heterogeneous and that the lower track students were lacking. Also, I feel that the fact that the students who did take part in the program were volunteers tended to affect the make up; since it appears that many of the white students either chose the class because they thought it would be easier or they tended to "political activist." The pilot class seemed to be quite negative in attitude compared to the homogeneous classes or to the heterogeneous classes this (1969-1970) year.

Materials: One of the most important problems facing the pilot project was to find adequate and suitable materials. This is a problem that we still have not completely solved. In the homogeneous classes a single text was used in each class with the text varying according to the ability of the class. This was not possible in the pilot group, and we were faced with the alternatives of (1) using one general text even though it might be too difficult or easy for many of the students and then supplementing it with other materials, or (2) using several basic texts with each student receiving the one

that best suited his reading ability. The decision of the teachers in the program, with the urging of our advisor from U. C., was to use the first alternative. I now feel that this was a mistake. In using the second method the material of the various texts must be tied together by the use of such things as worksheets, audio-visual materials, study groups and discussions.

It became apparent very early that there would have to be a greater dependence on the use of audio-visual materials. The pilot program was hanpered because, except for movies, these materials were lacking. However, it did focus our attention on the need.

Techniques: At first, the teachers participating in the heterogeneous pilot project felt that grouping in which the faster student could aid the slower student would be a useful teaching method. However, this technique proved of little value to the class, and in fact, grouping in general seemed to be of less value as a teaching technique in a heterogeneous class than in a homogeneous (especially upper track) class.

The technique of using students as tutors seemed to offer great promise at the beginning of the project. However, this did not work in actual practice. The faster students did not seem willing to give tutorial assistance nor did the slower students seem willing to receive it.

Summary: I have found the most success with a heterogeneous class by using several texts, with each student using the one that best fits his ability. A general class assignment using worksheets, audio-visual materials, discus-

sions, and group work ties the material in the various books together. In addition, various projects which may be done individually or by small groups are offered. These projects are optional and vary in difficulty. Thus, each student has an opportunity to reinforce his learning by doing additional work by choosing a project that corresponds with his interests and ability.

Some of the main points that are brought out in these statements are as follows:

1. Preparation of assignments--Minimal assignments were given to all students. Brighter students were encouraged to do extra assignments. The assignments were more varied than those normally found in a homogeneous class and there was a tendency to non-verbal kinds of assignments.

2. Reading materials--One text book was used. Apparently communication lines broke down by which teachers were urged to use as many different texts as necessary for their classes.

3. Discussion--Two teachers made greater use of discussion, while one made greater use of audio-visual materials.

4. Projects--There was a greater variety of projects which were conducted by all students.

5. Group work--Students were encouraged to work in groups and to help each other, though it appears that this part of the program was not too successful.

6. Use of teacher aide--One teacher used teacher aides to work with slow students and to assist brighter students in special projects.

These comments indicate that each teacher conducted his or her class the way that it was believed would be most successful or operational. As a result, the Track One students in the heterogeneous classes scored as well as the Track One students in homogeneous classes. Track Two students in the heterogeneous classes out-performed their counterparts in the Track Two homogeneous classes. In addition, the Track Three students trained in the heterogeneous classes obtained higher scores on the achievement tests than did the Track Three students trained in the homogeneous classes.

Summary

Statement of the Problem

Urban schools face the problem of giving equal opportunity to children with a variety of backgrounds and abilities. Due to the increase in pupil population in large school districts, educators have been forced to concentrate on supplying new school plants and have been unable to meet the challenge of finding and developing sound ways of educating children. In many school districts throughout the nation educators are in a quandry of what to do about tracking. In practice, if not in concept, the track system tends to discriminate against the disadvantaged child, particularly the Negro. While it is believed by many that tracking is one of the most efficient methods for teaching students with broad social and ability differences, the realities of the situation are that one of the unexpected consequences of tracking is that it has stigmatized the disadvantaged child, of whatever race, as being intellectually inferior and locked into an

educational track that further reinforces the disadvantages.

In most communities, schools are racially and socially heterogeneous and many so-called integrated schools are actually racially segregated within the school because of tracking. Educators have known for some time that these conditions damage the min ds and spirit of all children who attend them--the Negro, the white, the poor, and theaffluent--and block the attainment of the broader goals of democratic education, whether the segregation occurs by law or by fact. Knowing all this, it seems surprising that educators have not taken steps to abolish ability grouping and bring about real integration. However, a strong opinion prevails among teachers and administrators that ability grouping is necessary for effective instruction. Wheter it is actually necessary or desireable for effective instruction in eighth and ninth grade social studies classes is what this study has sought to determine.

Description of Grouping in Berkeley Schools

At the time of the study, the tracking practice in Berkeley Unified School District junior and senior high schools was to form four broad groups which were homogeneous with respect to a rather loosely defined combination of academic ability and achievement in the given subject. At the beginning of the seventh grade, the students were placed in groups by the counselors on the basis of the sixth grade teacher's recommendation and with the help of standardized tests (The Stanford Achievement Test and the Lorge-Thorndike Intelligence Test). After the first grading period there was considerable reassigning of the students to groups. In fact, all through a

student's school career he may have had his group assignment changed if there was evidence that he was misplaced.

In the junior high schools, English, history, science, and mathematics were grouped into four groups. In addition, classes in art, music, and foreign languages were divided, although along slightly different lines. Here is a qualitative summary of the seventh and eighth grade groups in social studies in the 1967-1968 school year. The top group was college prepatory and had about 35 percent of all the students. Its racial composition was overwhelmingly Caucasian and Oriental. Group Two spanned the middle of the range of abilities and was also college preparatory. It contained about 40 percent of the students. Its racial composition was about that of the whole student body, i.e., about 50 percent Caucasian, 40 percent Negro, and 10 percent Oriental. Group Three was not college prepatory. It consisted of students working considerably behind grade level and contained about 20 percent of the students. Its composition was 75 percent Negro. Group Four had about five percent of the school population and was overwhelmingly Negro. Special classes for the educationally handicapped and educationally mentally retarded were offered to those students who qualified.

Some Unintended Results of the Grouping System in the Berkeley Schools

1. De facto racial segregation. The most obvious unintended result of the grouping system in the Berkeley Unified School District (BUSD) is that it produced de facto segregated classrooms within the integrated secondary schools.

Racial stereotyping, social circumstances, and differences in educational background resulted in the top group being populated almost exclusively by Caucasians and Orientals and the bottom groups almost exclusively by Negroes.

2. Intellectual Isolation. The main idea behind grouping in the BUSD schools was the separation of the students according to intellectual or academic ability, so that the teacher could teach the separated groups more efficiently. Unfortunately, in addition to the possible benefits of this separation, there were some unintended undesireable side effects which resulted from such a structure. A student's classmates play a very important role in determining his intellectual progress. That this should be so makes a great deal of sense, since one of the causes of low achievement is lack of motivation and aspiration. It is unreasonable to expect a low-achieving student to pick up more favorable attitudes if his classroom contacts are limited to other students with low motivation and aspirations.

3. Cultural and economic isolation. It is well established that a child's performance in school (and on tests) is correlated to his cultural background and to the income level of his family. Generally speaking, middle and upper income families have parents whose educational level is higher than the parents of lower income families. Unfortunately, the educational background of the parents affects the achievement and attitudes of the children. The attitude with which a child approaches an IQ or aptitude test varies a great deal with his cultural background. Furthermore, the tests them-

selves take for granted a certain cultural background. As a result of these factors, the students who do well on the tests and who do well in school--and are consequently grouped together--tend to be from similar backgrounds. The same is true of those who do poorly. Thus, the grouped classes tend to be homogeneous with respect to cultural and economic factors as well as intellectual factors.

4. Lowering of aspirations. Another effect of the grouping system results from the fact that status, high or low, is attached to an individual on the basis of the groups he is placed in, or of which he is a member. The worst aspect of this is that a child in the lower group may suffer a crushing blow to his self-esteem by being labeled "lower group." His aspirations also tend to fall because not much is expected of him. He then accomplishes less than he is capable of accomplishing.

5. Stereotyping of students. Parents expect teachers to be sensitive to the individual needs of their children, and to teach them in such a way that these individual needs are met and individual talents developed. It is more difficult for a teacher to accomplish this aim when each child is remanded to him as part of a group which is presumed to have many characteristics in common, characteristics which may or may not hold for a particular child. It is difficult for a teacher to maintain his sensitivity to the individuality of a child if someone has already classified the child as part of a group. It is a temptation for the teacher to assume that the child's individual needs are being sufficiently met by the

grouping practice itself, thereby relieving the teacher of further responsibility for adapting his teaching to the individual child.

In most communities, the possible transition to heterogeneous classes will involve a great deal of careful planning and preparation. The detailed manner in which the transition is accomplished is the responsibility of the administration and the staff. This study was undertaken to help school administration decide whether homogeneous classes are indeed necessary for quality education in America's public schools.

Reasons for this study

The reasons for this study were put forth in the previous sections. In addition, several other objectives were of interest. They are:

1. To find out if students learn more in a heterogeneous classroom setting than in a homogeneous classroom setting.

2. To help administrators and teachers structure a quality educational program attuned to the demands of best educational practice.

3. To help create an educational program responsive to the needs of the Berkeley community.

4. To show the community that the BUSD decision to integrate pupils is a responsible, well-thought-out, statistically valid, non-emotional progressive act.

This report focusses on the first of these objectives.

Prodedure used to define the experimental groups of the study

At the end of the spring semester of 1967, a questionnaire was distributed to all parents and guardians of students assigned to the state required classes in eighth grade social studies. In the questionnaire, parents were informed of the intent of the school district to perform the study described in this report. Each parent was invited to participate in the study by volunteering their children for assignment to a heterogeneous social studies class for the upcoming school year.

On the basis of the parent response the entire entering eighth grade was divided into a Volunteer Group, comprising students whose parents expressed a willingness for the participation of their children in the experimental heterogeneous classes, a Non-Volunteer Group, comprising students whose parents expressed an opposition to participation in the experimental classes, and a Non-Response Group, comprising students whose parents made no reply to the question of whether they approved of their children's participation or opposed it.

From the Volunteer Group, the students for the heterogeneous classes were randomly selected from social studies Track One, Two, and Three classes. These students were randomly and evenly distributed among four experimental classes, in a proportion which represented the proportion of each group in the eighth grade school population as a whole; that is, for each experimental class, 10 students were from Track One, 11 students were from Track Two, and seven students were from Track Three. The balance of the Volunteer Group were evenly and randomly distributed into their regular tracked

classes. This last Volunteer Group can be considered as a control class designed to eliminate any bias inherent in the act of volunteering.

The Non-Volunteer Group was evenly and randomly distributed across the projected tracked classes and the Non-Response Group, also, was evenly and randomly distributed across the ordinary tracked classes. Efforts were made to maintain a proportionate balance representative of the whole grade between boys and girls in each distribution, though it was obvious that this was not possible in groups where there were normally more boys than girls or where the opposite condition existed. In any case, the proportion of boys to girls in each class was approximately equivalent to the normal distribution of these groups in the total eighth grade class.

The number of classes available for the first years' study was as follows:

Group I	6 classes of size 32
Group II	7 classes of size 32
Group III	3 classes of size 25
Experimental Heterogeneous	4 classes of size 28
Group IV	3 classes of size 17
Honors	1 class of size 20

As this breakdown and distribution of students suggests, each heterogeneous class was composed of approximately 10 Track One students, 11 Track Two students and seven Track Three students. This means that each heterogeneous class has a fairly large group of high ability students since it must be recalled that many of the Track Two students are college potential as are

most of the Track One students. In the interpretation of results, this must be taken into consideration.

As far as was practicable, each of the teachers of social studies was assigned to a full range of classes: heterogeneous class and homogeneous classes from Tracks One, Two, and Three.

Statistical procedures used to process data

The object of the analysis is to compare and ascertain differences that may be manifested between and within classes on the criterion measures. Analyses of variance are used to determine if statistically significant differences exist at the end of the experimental terms and school years. Appropost hoc comparisons and confidence intervals using Scheffé's method of multiple comparisons are used to assess the value of significant findings. The comparisons of interest are between the students of each group who are in the experimental heterogeneous classes and those of the same group who are in the ordinary homogeneous classes. Comparisons are also made to determine if students whose parents have indicated a willingness for them to participate in the experiment are different on the criterion measures from those whose parents have opposed participation or who have not responded. By distributing the classes evenly among the teachers, as far as this was possible, the effects of teacher differences were hopefully controlled in the analysis.

For the analysis of attitude data, factor analyses were performed to create scale values. These scale values were then used in analysis of variance designs to study the

basic hypotheses of the study. Stated in positive form, these hypotheses are:

1. Among the various ability grouped children placed in the heterogeneous classes Track One children will do as well as the Track One children in the homogeneous classes.

2. The Track Two children placed in the heterogeneous classes will do significantly better than those Track Two children who are placed in the homogeneous classes.

3. The Track Three children placed in the heterogeneous classes will do significantly better than those Track Three children who are placed in the homogeneous classes.

As a starting point, it seemed reasonable to assume that students whose parents were concerned about their children's performance and/or about the possible effects of heterogeneous grouping upon school learning and social development would also be subject to differences in either home environment or home attitude. Therefore, perhaps the most important comparisons for the evaluation are the mean differences between the Volunteers in the experimental classes and the non-selected Volunteers who had been distributed across the regular homogeneous classes. It is believed that the comparisons between these two groups holds constant any differences due to the forces operating on students and parents which resulted in the event "volunteering for the study." Thus, these comparisons are given special attention in the analysis of the data.

Initial differences in school achievement and ability of the experimental groups

Results from the regular testing of the eighth grade

students on the Sequential Tests of Educational Progress (STEP) and the School and College Ability Test (SCAT), administered in the fall of 1967, were used to determine if differences in ability existed between the three groups or kinds of students within each track. The F-ratios of the analysis of variance indicated that reliable differences existed only within the Track Two groups. The heterogeneous volunteer group of Track Two did better on the STEP and on the verbal subsection of the SCAT. The scores were standardized so that the mean for the entire group of students in the study was equal to 50. In this form the track means and the group means within tracks were as shown in Table I.

While the difference in verbal scores between the heterogeneous and homogeneous volunteers in Track Two is given by $\hat{\Delta} = 50.1 - 46.0 = 4.1$ standard score units, it must be looked upon as a random or chance deviation since both groups of students were generated at random from the one pool of volunteer students, for which the mean score of 48.0 is identical to the mean score of the non-volunteers. Also, the difference is not significant by Scheffé's Theorem, suggesting that the groups of students within the individual tracks were equivalent with respect to verbal ability at the beginning of the study.

No significant differences were found for any of the "within" groups on the SCAT quantitative section or on the total SCAT score which represents an estimate of total ability. Thus, it can be concluded that with respect to quantitative ability the students within the experimental groups for each

Table I. Mean Scores on the Pre-test Achievement Measures for the 1967-1968 School Year.

Experimental Group of the Study	Achievement Measures STEP	SCAT Verbal	SCAT Quant	SCAT Total	U.S. Constitution
Track One Students					
Heterogeneous Volunteers	55.4	56.0	54.9	55.8	54.1
Homogeneous Volunteers	56.5	57.3	56.6	57.7	54.9
Homogeneous Non-volunteers	57.0	57.8	55.7	58.1	55.9
Track Two Students					
Heterogeneous Volunteers	50.6	50.1	50.0	48.7	48.8
Homogeneous Volunteers	47.6	46.0	48.3	46.2	48.7
Homogeneous Non-volunteers	46.9	48.0	49.2	47.8	46.5
Track Three Students					
Heterogeneous Volunteers	41.3	39.7	42.5	39.5	45.0
Homogeneous Volunteers	41.3	35.1	40.3	35.5	46.5
Homogeneous Non-volunteers	37.5	37.3	41.8	37.5	42.5
F-ratios for the Analysis of Variance Tests of Equal Mean Values					
Between Tracks	17.6*	252.7*	83.6*	270.2*	68.0*
Groups in Track One	.6	.9	.3	1.6	.6
Groups in Track Two	3.5*	3.1*	.4	2.0	1.4
Groups in Track Three	2.5	1.6	.2	1.2	.7

*Differences are statistically significant at the 5 percent level of significance.

track are equivalent. In this sense, volunteering for assignment to the heterogeneous class was not based on quantitative ability, as measured by SCAT.

Performance in School Following the First and Second Semester of the 1967-1968 School Year

The overall grade point averages at the end of the first semester indicated that there were no differences between the groups within each track. It would seem that whatever differences may have existed within the Track Two groups, their overall performance as rated by their teachers resulted in no difference. The mean grade point averages, standardized on the total group are as shown in Table II. Thus, with respect to grade point average, the groups within a track are equivalent following one semester of training. In this sense, heterogeneous grouping in social studies does not affect overall school performance.

During the second semester, the Cooperative Social Studies Test was administered to all students in social studies. This test was used to indicate possible differences in achievement that might result from the formation of the experimental classes. The teachers involved indicated that part of this test could be assumed to tap elements of the curriculum which were taught, but that part of the test contained questions about materials which were not taught but might be peripherally learned. For this reason, the items on the test were divided into two sets of items. Set One contained the questions that seemed, to the teachers, to be testing what the students should have learned; Set Two contained the questions which seemed, to

Table II. Mean Scores on the Post-test Achievement Measures for the 1967-1968 School Year and GPA After First and Second Semesters.

Experimental Group of the Study	Achievement Measures: GPA First Sem.	GPA Second Sem.	CSST** (Set one)	CSST (Set two)	U.S. Constitution
Track One Students					
Heterogeneous Volunteers	54.1	53.7	52.7	52.0	54.1
Homogeneous Volunteers	56.9	56.3	56.4	56.0	55.5
Homogeneous Non-volunteers	56.8	56.5	57.4	56.0	56.3
Track Two Students					
Heterogeneous Volunteers	47.5	47.6	46.5	46.9	47.6
Homogeneous Volunteers	46.5	45.9	47.2	47.1	45.1
Homogeneous Non-volunteers	48.4	48.3	48.6	48.4	48.8
Track Three Students					
Heterogeneous Volunteers	42.8	43.6	40.2	44.3	42.9
Homogeneous Volunteers	40.1	46.0	43.0	42.3	40.3
Homogeneous Non-volunteers	39.2	40.6	39.7	42.3	37.7
F-ratios for the Analysis of Variance Tests of Equal Mean Values					
Between Tracks	134.2*	109.9*	147.3*	73.1*	110.2*
Groups in Track One	1.7	1.7	4.5*	3.3*	1.0
Groups in Track Two	.9	1.4	1.2	.7	1.5
Groups in Track Three	1.8	2.3	.6	.5	3.2*

*Differences are statistically significant at the five percent level of significance.

**Cooperative Social Studies Test.

the teachers, to be testing an area which was not directly taught, but which might have been peripherally learned.

The resulting statistics indicate that the Track One students in the heterogeneous classes did less well on both versions of this test, and that the Track Two and Track Three students in the heterogeneous classes did as well, but not better, as the other students in their respective tracks. Thus, none of the three hypotheses of the study were supported as stated. However, it was noted that the poorer performance of the heterogeneous Track One students was not statistically different from the Track One homogeneous volunteers, their appropriate control group.

A second criterion measure of social studies achievement was devised using the test on the Constitution required for the eighth grade. Previous to the teaching of the unit specifically preparing the students for the test, a selection of questions supplied by the textbook publishers was given to all students in the study to see if there were initial differences between the groups within the tracks. Subsequent to the teaching of the unit on the Constitution, a test, again using questions supplied by the textbook publishers, was given to indicate differences in achievement on the unit.

The results indicated that on the pre test there were no significant differences within the tracks. The results on the post test indicated that the Track Three volunteers in the heterogeneous classes did get higher scores than the other Track Three students. This may indicate that heterogeneous classes were taught with greater reliance on the textbook as

a guide, or alternatively, more concrete questions on facts from the text tapped differences in the Track Three students which the Cooperative Social Studies Test did not. The Track One and Track Two students did as well as, but not better than, the other students who were taught in the homogeneous classes. Thus, two of the three stated hypotheses were supported for this particular test.

The overall grade point average for the end of the school year indicated that there were no differences between the groups within the various tracks.

Attitudes Towards Classwork and Other Students in Social Studies

A series of attitude questionnaires was administered during the second semester. The items in this questionnaire were aimed at assessing attitudes toward self, school, social studies classes, and other relevant areas. The responses to these questionnaires were submitted to a factor analysis, whereupon those questions which did not appear, on the basis of the preliminary run, as discriminating questions were eliminated from the battery; finally, the remaining questions were factor analyzed a second time. The measures used for comparison were factor scores based upon the rotated factors from the second analysis. These scores were standardized across the total group so that the average is equal to 50, with a standard deviation of 10. Students who scored high on the dimension of "self" reported themselves as being pleasant, happy, valuable, beautiful, etc., whereas students with poor self-images saw themselves as unpleasant, sad, worthless, ugly, etc.

In the evaluation of "self", reliable differences existed between the tracks and within Track Three. While various interpretations may be offered to explain the results of the analysis of variance on the factor scores, a description of the differences might be summarized as follows: On the whole, Track One students are more self-critical, perhaps because they are more sensitive to differences between their conception of "ideal" behavior and their own behavior and because the potential and real activities of their peers are of a higher standard. The more interesting result of the analysis occurred in the three groups comprising the Track Three students. Those volunteers who were assigned to the experimental classes exhibited the same tendency toward self-criticism. This may represent the acquisition of a more realistic interpretation of their own behavior or it may be simply the acquisition of new attitudes from their heterogeneous groups since this measure was administered in the context of the social studies classes.

The means of the nine groups are shown in Table III. For the Track Three students the mean difference for heterogeneous versus homogeneous students is given by $\hat{\Delta} = 47.3 - 55.8 = -8.5$ standard units. If the reduced self-image scores, similar to the Track One means of 48.3, represents a true or more honest appraisal by these students, then it would appear that heterogeneous grouping is indeed a valuable adjunct to determining self-worth and self-image. Students who are complacent and satisfied will not try to improve their status and standing whereas a student who believes there is room for improvement will try harder and seek to improve. Such an improvement was

Table III. Mean Scores on the Attitude Measures Following the 1967-1968 School Year.

Experimental Group of the Study	Attitude Measures of: Self	Classmates this year	Social Studies last year	Social Studies this year	Classmates in Social Studies last year	Classmates in Social Studies this year
Track One Students						
Heterogeneous Volunteers	47.7	49.8	49.3	46.2	51.8	50.6
Homogeneous Volunteers	45.3	49.9	47.4	51.0	51.0	50.1
Homogeneous Non-volunteers	49.2	50.9	51.3	51.6	54.0	54.0
Track Two Students						
Heterogeneous Volunteers	51.1	49.7	50.7	46.3	48.0	46.8
Homogeneous Volunteers	49.6	45.9	51.4	46.8	50.0	47.5
Homogeneous Non-volunteers	51.2	50.6	48.6	50.9	48.4	50.5
Track Three Students						
Heterogeneous Volunteers	47.3	50.7	50.1	44.7	45.3	40.2
Homogeneous Volunteers	58.0	45.9	54.5	55.0	48.9	49.0
Homogeneous Non-volunteers	55.5	49.2	51.3	52.7	44.2	45.6
F-ratios for the Analysis of Variance Tests of Equal Mean Values						
Between Tracks	6.3*	.3	.9	.4	18.7*	18.3*
Groups in Track One	1.7	.2	2.0	4.2*	1.4	2.7
Groups in Track Two	.3	1.8	1.3	4.3*	.4	2.3
Groups in Track Three	4.5*	.4	.9	5.2*	.5	2.6

*Differences are statistically significant at the five percent level of significance.

noted for the Constitution post test. In this sense, a more honest self-appraisal with a reduced self-image score was accompanied with increased school learning as measured by scores on the Constitution post test.

A second factor analysis was performed on the adjective pairs that were used to define the way students felt about the other students in the school. Students with positive feelings about the other students report them as being nice, fragrant, clean, sweet, fair, etc., while students with unfavorable feelings report their fellow classmates as awful, foul, dirty, sour, unfair, etc. This evaluation measure of other kids in school is considerably unlike evaluation measures related to self. While the attitude measure of self is an evaluate factor concerning worthiness and value of a person, the attitude measure of the other students is a factor describing student behavior as perceived by students making the evaluation. Students with favorable attitudes on this dimension report their peers as nice, clean, and fair, while at the opposite pole students are reported as awful, dirty, and unfair. The mean scores for the nine study groups are presented in Table III. As can be seen, none of the mean differences are statistically significant, whether they are made between the tracks or within the tracks.

Students were asked to recall their attitudes toward their classmates in their seventh grade social studies classes. The mean scores for attitude toward classmates in social studies last year are summarized in Table III. In this case, there are large differences between the tracks but no dif-

ferences within the tracks. Students in Track One tended to report that the kids in World History and Geography were as smart as them, they hardly ever felt impatient with the slow kids, they felt at ease with the other kids, and they respected the other kids as individuals. On the other hand, Track Three students tended to report the opposite about the kids that were in their World History and Geography classes. Since these classes were tracked, this finding is of some interest in that Track Three students did not hold favorable attitudes towards other Track Three students in the seventh grade social studies class.

On questions aimed at attitudes towards classwork and assignments this year, the Track One heterogeneous group liked this aspect less than the other Track One groups liked theirs. In Track Two, both the Selected and Non-selected Volunteer groups liked their classwork and assignments less than the Non-volunteer group. In Track Three, the heterogeneous group liked their classwork and assignments less well than either of the other two Track Three groups.

Mean scores for attitudes towards classwork and assignments in social studies during the 1967-1968 experimental year are shown in Table III. In this case, each within track source of variance is statistically significant. In each case, it is the heterogeneous volunteers that report the least satisfaction with the assignments and classwork. The mean differences for the three tracks of students are given by:

$$\hat{\Delta}_1 = 51.5 - 46.2 = 5.3$$

$$\hat{\Delta}_2 = 50.7 - 46.3 = 4.4$$

$$\hat{\Delta}_3 = 52.9 - 44.7 = 8.2$$

suggesting that the dissatisfaction increases as the track number increases. This is somewhat surprising when it is recalled that the Track Three students showed considerably more learning on theConstitution post test.

The mean scores of attitudes towards classmates in social studies during the experimental year are also shown in Table III. None of the differences within the individual tracks are statistically significant. Only the difference between the tracks is significant, but this difference parallels that noticed for the corresponding items relating to last year.

Attitudes of the Adults at Home to the School Environment as Perceived by the Students

A series of questions designed to elicit attitudes on the adults at home as projected by the students yielded some differences. The questions were worded so that the students were to answer on the basis of what they knew or believed about the attitudes of their parents or guardians. It was not expected that this would necessarily reflect the true attitudes of the adults in the family.

Track One students as a whole said the adults in their home approved of the behavior of the student body and the educational quality of the school. Track Two students reflected a less approving attitude and the Track Three students reflected the least approving attitude. There were no differences on this variable for any of the within track group

comparisons.

On the questions regarding grouping or tracking in the school, the Track One volunteers assigned to the heterogeneous classes said their parents disapproved of tracking, the Track One volunteers relegated to the regular classes said their parents disapproved of tracking, (but not to the degree of the experimental group), while the Track One non-volunteers said that their parents approved of ability grouping in school. There were no reliable differences between the tracks or within Tracks Two and Three.

Questions regarding attitudes towards the administration and faculty at the school indicated that no differences existed across the groups as the factor scores were uniform across the nine groups.

Summary Conclusions for the 1967-1968 School Year

The general results for the 1967-1968 school year are not impressive. There were indications that the Track One students did not achieve as well under heterogeneous grouping by about .3 to .4 of a standard deviation. The teachers involved with the heterogeneous classes reported difficulties in stimulating the Track One students in the experimental classes to do well relative to their own ability rather than to do well relative to the class average. The teachers felt that they did not have enough time to prepare adequately before the advent of the first year's study to stimulate all the higher ability students in the experimental classes, but in any case it should be noted that the difference between heterogeneous and homogeneous trained students is not significant.

The higher scores of the Track Three students in the experimental classes on the Constitution post test is some indication that Track Three students can improve in a more heterogeneous group, at least when the questions asked are of the more concrete textbook, factual nature.

The main conclusions that can be drawn from the first year's study are:

1. Heterogeneous trained students perform as well as their counterparts in the homogeneous classes on the relevant items of the Cooperative Social Studies Test.

2. Students trained in heterogeneous classes did as well on incidental learning for the Cooperative Social Studies Test as did their counterparts trained in homogeneous classes.

3. Heterogeneous volunteers in Track Three showed higher scores on the United States Constitution post test.

4. Heterogeneous volunteers in Track Three showed a self-image that was more critical and more in line with the students in Track One.

5. Students in the heterogeneous classes were less satisfied with their class assignments during the experimental year.

6. It is quite clear that students who were in the heterogeneous volunteer classes or who did volunteer but were assigned to the homogeneous classes reported that their parents did not look with favor upon ability groupings and tracking.

The Experimental Groups for the 1968-1969 School Year

During the second year of the study, students transferred to the freshman campus of the senior high school. In

addition, they were joined by students from the remaining junior high school of the district. These new students were given an opportunity to join the study if their parents so desired. In addition, the students in the study of the previous year were given the opportunity to continue in the new classes.

From the volunteers, heterogeneous classes were created by selecting students at random. The remaining students were assigned to homogeneous classes.

Initial Differences in School Achievement and Ability of the Experimental Groups of the 1968-1969 School Year

At the beginning of the second year's study, all students in the school were given the Differential Aptitude Test. The corresponding mean standardized scores are shown in Table IV. Except for the differences between the tracks, differences between comparable groups of students are not significant.

Track One students average 58.2, or almost one standard deviation above the mean score for the 904 students tested. The Track Two students averaged 46.4, while the Track Three students averaged 40.2. These large differences in verbal aptitude between the tracks is not unexpected. Within the Track One students, the heterogeneous volunteers scored 57.0, the homogeneous volunteers scored 58.8, and the homogeneous non-volunteers scored 58.2. As expected, these three means are not statistically significant from one another. For the heterogeneous volunteers, the continuing students scored 55.3 while the new heterogeneous volunteers scored 58.5. In

Table IV. Mean Scores on the Pre Test Achievement Measures for the 1968-1969 School Year.

Experimental Group of the Study	DAT Verbal	DAT Numeric	DAT Spatial Relations
Track One Students			
Heterogeneous Volunteers	57.0	57.1	56.4
Homogeneous Volunteers	58.8	57.9	57.6
Homogeneous Non-volunteers	58.2	57.2	56.4
Track Two Students			
Heterogeneous Volunteers	48.6	48.7	49.1
Homogeneous Volunteers	46.1	47.6	49.1
Homogeneous Non-volunteers	46.2	46.8	48.1
Track Three Students			
Heterogeneous Volunteers	41.7	41.9	41.5
Homogeneous Volunteers	40.8	42.0	39.1
Homogeneous Non-volunteers	40.1	40.4	39.4
F-ratios for the Analysis of Variance Tests of Equal Mean Values			
Between Tracks	476.4*	421.1*	286.1*
Groups in Track One	1.7	.1	.5
Groups in Track Two	1.7	.9	.4
Groups in Track Three	.5	.4	.5

*Differences are statistically significant at the five percent level of significance.

this case, this difference is not significant. Similar statements can be made about Track Two and Tract Three students. Thus, with respect to verbal aptitude as measured by this test, there are no differences between the continuing heterogeneous volunteers and the new heterogeneous volunteers coming from the other junior high school in the school district.

Thus, it can be concluded that differences in verbal ability of the students in the three tracks are statistically significant, but within the tracks there are no statistically significant differences between the students placed in the heterogeneous or homogeneous classes. Also, the new students coming into the study have verbal ability scores that are essentially the same as those of the students in the previous year's study.

The mean scores on the numeric part of the Differential Aptitude Test are shown in Table IV. None of the comparisons are statistically significant. Mean standardized scores for the spatial relations items of the Differential Aptitude Test for the various experimental groups are also shown in Table IV. None of the differences are significant.

It is concluded that the students in the various conditions of the study are well matched with respect to verbal ability, numerical ability, and spatial relations as measured by the Differential Aptitude Test.

Performance in Social Studies Following the Second Year's Study

In mid-May of 1969, all students in the school were given a social studies test that was designed by the teachers acting as a collective body. This included teachers who taught

students in heterogeneous classes as well as teachers who taught students in the homogeneous classes. The objectives of the test were to tap the knowledge and skills that the teachers thought should have been covered in the classes and acquired by the students to show that they had successfully learned the material of the curriculum. The mean scores for the test are presented in Table V.

For the Track One students, the F-ratio for the comparison between the heterogeneous volunteers, the homogeneous volunteers, and the homogeneous non-volunteers is given by $F = 1.06$ which is not significant. The corresponding mean scores are 57.9, 56.9, and 56.0. This is somewhat different from the results of the first year's study in which it was noted that the Track One students in heterogeneous classes did not do as well as Track One students in homogeneous classes, even though their performance was not statistically different from the average Track One student. The teachers in the first year reported that motivating the Track One students was something of a problem. This problem seems not to have been too important for the second year's study. Since the teachers of the second year's study knew of this difficulty, they made extra efforts to motivate these children. On this basis, it is reasonable to conclude that performance of Track One students is unrelated to the ability make-up of the classroom, provided that teachers take the effort to motivate Track One students to learn and act in a fashion commensurate with their favorable intellectual endowment. Thus, the first hypothesis of the study has been supported.

Table V. Mean Scores on the Post Test Achievement Measures for the 1968-1969 School Year.

Experimental Group of the Study	Achievement Measures	
	Teacher-made Test	CSST**
Track One Students		
Heterogeneous Volunteers	57.9	57.2
Homogeneous Volunteers	56.9	54.3
Homogeneous Non-volunteers	56.0	56.5
Track Two Students		
Heterogeneous Volunteers	52.8	51.8
Homogeneous Volunteers	47.3	47.0
Homogeneous Non-volunteers	47.6	47.6
Track Three Students		
Heterogeneous Volunteers	42.3	38.1
Homogeneous Volunteers	38.2	33.6
Homogeneous Non-volunteers	36.1	36.1

F-ratios for the Analysis of Variance Tests of Equal Mean Values

Between Tracks	323.5*	1730.9*
Groups in Track One	1.1	.3
Groups in Track Two	7.4*	2.9***
Groups in Track Three	6.0*	3.1***

*Differences are statistically significant at the five percent level of significance.

**Cooperative Social Studies Test

***Differences are statistically significant at the ten percent level of significance.

For Track Two students the findings are quite different. In this case, the F-ratio for the comparisons of heterogeneous volunteers, homogeneous volunteers, and homogeneous non-volunteers is statistically significant since F = 7.43. The mean scores for these three groups are 52.8, 47.3, and 47.6, respectively. Thus, it can be concluded that for Track Two students, training in a heterogeneous or homogeneous class does make a difference. In terms of standardized units, students trained in a heterogeneous class outscore their Track Two counterparts trained in homogeneous classes by .5 of a standard deviation. Thus, the second hypothesis of the study has been supported.

For Track Three students the differences involving the heterogeneous volunteers, the homogeneous volunteers, and the homogeneous non-volunteers are significant. The corresponding mean scores are given by 42.3, 38.2, and 36.1 with the F-ratio given by F = 5.96. Thus, it is concluded that the Track Three students trained in heterogeneous classes outperform the Track Three students trained in homogeneous classes by more than .6 of a standard deviation. For educational variables, differences of this magnitude are indicative of a major statistical finding. In addition, it is concluded that the third hypothesis of the study has been supported.

For the Cooperative Social Studies Test, the results are not so impressive even though the differences are in the same direction as those noted for the teacher-made test. In any case, none of the differences are statistically significant at the .05 level. Even though the results are not statistically

significant at the .05 level, exactly the same findings as noted for the teacher-made test are significant at the .10 level. If the sample sizes for the Cooperative Social Studies Test had been the same as those for the teacher-made test, the results would probably have been statistically significant at the .05 level.

It should be noted that the three hypotheses of the study are referred to as directional hypotheses by statisticians. For hypotheses of this nature, the classical F-test is not the best test to use since it is insensitive to predicted directions and outcomes since it is an omnibus test of significance. If the three hypotheses of the study are given directional tests, then all three hypotheses of the study are supported at the .05 level.

Attitudes Towards Classwork, Self, and Other Students in Social Studies During the 1968-1969 School Year

During the second year of the study, students were given an attitude inventory at the beginning of the school year. This inventory was considerably shorter than the one given during the first year mainly because the first year's questionnaire was revised and improved by discarding nondiscriminating items. Since the results of the pre test parallel the post test of the first year, the results are not presented and discussed. Instead, the post test results for the second year are examined. In making this evaluation, it should be recalled that a large number of students had completed two full years in heterogeneous classes. If being placed in a heterogeneous class induces fundamental changes in attitude

then there was sufficient time for such changes to have occurred. As will be noted, few changes did occur, suggesting that ability mixing does not have a strong impact on student's attitudes.

The mean scores of the various study groups of the study are shown in Table VI. As expected, the evaluation of self varies across the three tracks; F = 5.40 represents a significant difference at the .05 level. The mean standardized scores for the three tracks are 48.7, 50.7, and 52.0. Thus, similar to the first year's study, the Track One students are more critical of themselves than are the Track Three students. However, the degree of self-criticism is not as variable as the previous year where the mean scores were given by 48.3, 50.9, and 54.2.

The movement to a uniform self-evaluation across tracks is somewhat surprising, especially when it is noted that most of the classes in other academic subjects in the second year study were homogeneous. Among the Track Two students, the differences in attitude are statistically significant. The homogeneous volunteers scored 44.6 while the remaining students in the other two groups scored 51.2 and 51.3. Similar to the finding for the Track One students, none of the differences between the groups for the Track Three students are significant.

The mean scores for attitudes towards other kids in this school are shown in Table VI. For the differences between the tracks, the F-value is given by F = 5.72, with the mean scores for the three tracks given by 51.6, 49.0, and 49.4. While the differences are significant, they are so small as to

Table VI. Attitudes of Students Following the 1968-1969 School Year.

Experimental Groups of the Study	Attitude Measures Towards Self	Social Studies	Other Kids
Track One Students			
Heterogeneous Volunteers	50.1	48.0	53.0
Homogeneous Volunteers	46.6	47.5	54.3
Homogeneous Non-volunteers	48.9	48.2	51.1
Track Two Students			
Heterogeneous Volunteers	51.2	52.0	49.5
Homogeneous Volunteers	44.6	48.7	49.6
Homogeneous Non-volunteers	51.3	49.4	48.9
Track Three Students			
Heterogeneous Volunteers	48.9	44.0	48.9
Homogeneous Volunteers	55.9	45.4	55.9
Homogeneous Non-volunteers	48.7	47.9	48.7
F-ratios for the Analysis of Variance Tests of Equal Mean Values			
Between Tracks	5.4*	9.7*	5.7*
Groups in Track One	1.1	.1	2.1
Groups in Track Two	4.6*	1.0	.1
Groups in Track Three	2.2	1.2	2.3

*Differences are statistically significant at the five percent level of significance.

be meaningless.

Mean scores for attitude towards classwork and assignments in social studies during the second year's investigation are shown in Table VI. Outside of the differences between the tracks, none of the comparisons involving heterogeneous versus homogeneous classes are significant. Thus, assignment in heterogeneous or homogeneous classes has no impact upon the way students feel about classwork or assignments during the second year of the study with respect to social studies. As expected, the students in Track Three show the most dissatisfaction with the assignments and classwork in the social studies classes; however, the dissatisfaction is not very strong.

Concerning what the parents at home think, two factors were generated by the factor analysis. Concerning educational quality, there are no differences between the three tracks of students and within the tracks there are no statistically significant differences. However, for the attitudes related to ability groupings there are significant differences between the groups with F = 14.92. According to the reported responses, students of Track One report their parents are most in favor of ability tracking and grouping, while parents of students in Track Three classes are reported as being opposed. The mean scores are 47.4, 51.0, and 53.4.

Summary Conclusions for the 1968-1969 School Year

1. Students in the various experimental conditions of the study were well matched at the beginning of the school year with respect to verbal ability, numerical ability, and

spatial relations as measured by the Differential Aptitude Test. While there were large differences between the students in the three different tracks, the differences between the groups within the tracks were not significant.

2. For the teacher-made test, the performance for Track One students was independent of the ability group spread within the classes. Students placed in heterogeneous classes did as well as students placed in homogeneous classes. The mean standardized scores for the heterogeneous volunteers was 57.9, while the mean scores for the homogeneous volunteers and non-volunteers were given by 56.9 and 56.0, respectively.

3. For the teacher-made test, performance of the Track Two students was related to the class in which their training occurred. Students placed in the heterogeneous classes out-performed their homogeneous counterparts by 5.2 standardized units. The mean score for the heterogeneous trained students was 52.8, while the mean score for the homogeneous trained students was 47.6.

4. For the teacher-made test, the performance of the Track Three students was related to the ability make-up of the class in which instruction took place. The mean score for the students assigned to the heterogeneous classes was 42.3 while the mean score for the students trained in the homogeneous classes was 36.2.

5. For the Cooperative Social Studies Test, none of the major differences were significant at the .05 level, though all are in the direction noted for the teacher-made test. Since the sample sizes for this test were considerably smaller,

it is quite possible that if a larger sample had been available, the decisions made with respect to the teacher-made test would also be made with respect to this test.

6. As was noted for the first year study, Track One students are more critical of themselves than are the students of Track Two and Track Three.

7. For the self-evaluation measures of the second year study, the variability is less than the first year study.

8. When students were asked to report how they feel about other kids in their social studies classes, it was found that their attitudes were not affected by training in heterogeneous or homogeneous classes.

9. When asked to report how they felt about the class assignements and school work, it was found that these attitudes were unrelated to training in heterogeneous or homogeneous classes.

10. Students in Track One report that their parents favor ability grouping while students in Track Three report that their parents are not in favor of ability groupings.

Discussion

The major findings of this study are that student attitude is not affected to any pronounced degree by training in a heterogeneous or homogeneous class. Attitudes of an individual towards self, classmates, and school seem to have little relationship to the ability of the remaining students in the class. However, the same conclusions do not apply to individual school achievement. In this study, the performance of Track One students was not affected by training in hetero-

geneous or homogeneous social studies classes, at least as measured by the tests of this investigation. On the other hand, students of Track Two and Three performed at a higher level in a class that was heterogeneous with respect to ability. Possible reasons for this finding are numerous. The most frequently reported reasons for this finding are that low achieving students are exposed to models that they can try to emulate and that the opportunity to interact with high achieving students increases the richness of the environment because of the greater variety of learning situations and experiences that become available to poor achieving students. Both of these explanations were discussed by Coleman.

According to Coleman (pp. 302-304), "Attributes of other students account for far more variation in the achievement of minority group children than do any attributes of school facilities and slightly more than do attributes of staff.

"In general, as the educational aspirations and background of fellow students increases, the achievement of minority group children increases... The results suggest ... that the environment provided by the student body is asymmetric in its effects, that it has greatest effect on those from educationally deficient backgrounds." In addition, as reported (pp. 307-308), "as the proportion white in a school increases, the achievement of students in each racial group increases."

Since Track Three students tend to be Negro and Track One students tend to be white, the results of this investigation suggest that Coleman's comments about schools have a direct parallel to classrooms. The academic achieve-

ment of Track Three and Track Two students was higher in the mixed classrooms, while Track One students were essentially unaffected. Both of these classroom findings are in agreement with Coleman's statements about schools.

Another explanation for these findings might be related to the experience, know-how, interest, and quality of the teachers and what they did in the classrooms while teaching the students with variable abilities. At the beginning of the study, it was decided to give the teachers free reign over their classes and as such, they were told to conduct their class in what they thought the best way to teach their pupils. They were supplied with the regular text books for Tracks One, Two, and Three social studies classes. Students were to be permitted to choose the level of text they wanted to use. This practice was followed during the first, but not the second year of the study. While the curriculums by text across the tracks were essentially identical, their presentations were at different levels of complexity, depth, vocabulary, and interest. In addition to multiple texts, the usual maps, films, globes, related texts, newspapers, magazines, or other educational hardware were available to the teachers for class use, if so desired. As this suggests, the experimental procedures used in the heterogeneous as well as those used in the homogeneous classes were unstructured and left to the interests and imaginations of the teachers. This procedure was encouraged and followed because it is believed that this is the way teaching would normally take place in a typical classroom. This means that teachers were free to decide for themselves the manner in which

they wanted to conduct their class. What biassing effect this could have had on the outcomes of the study are unclear. If it accounts for a "true" improvement in student achievement, then mixing the tracks has been effective. However, if it only produced a Hawthorne effect, then the outcomes are not so striking. Only future research can show which was actually the case.

Because of the unstructured experimental condition, the immediate question is what did the teachers do in the experimental classes that was different from what they did in the regular tracked classes? Some of the main differences for the first year are:

1. Preparation of assignments--Teachers reproduced materials for students which were designed for two reading levels. The students were allowed to choose either assignment. The teachers indicated that many students did both assignments.

2. Reading materials--Besides the regular textbook, supplementary books for students of varying abilities were used.

3. Discussions--There apparently was more class discussion in the pilot classes than in the Track Three and Track Four classes and also there was greater student preparation.

4. Projects--Various suggestions for projects were made to accommodate the students' abilities and interests. (Scrapbooks, written reports, etc.)

5. Group work--Students were encouraged to work in

groups and to help each other.

6. Tests--These were designed for students of all abilities. Questions were both required and optional, with extra credit given for the latter. More difficult questions were used to challenge the more able students.

As these comments suggest, the teachers' activities in the heterogeneous classes were not much different from their activities in a homogeneous class, but what seemed to differentiate the heterogeneous class from the homogeneous class was the interaction that took place between the students. As suggested by one of the teachers, the emphasis on methods was different, with more cooperation and helpfulness between the students. While this may not have been too effective for improving school learning during the first year of the study, one might suspect it was for the Track Two and Track Three students during the second year of the study.

As to what teachers did during the second year of the study in the classrooms is not clear and where information is available, it seems that the activities were not uniform. Judging from the comments made by the teachers of the first year's study and the improvement in achievement scores of the students during the second year, one might assume that more ideal teaching circumstances were encountered during the second year of the study. The comments of three of the involved teachers give partial support for this supposition. Some of the main points of difference are:

1. Preparation of assignments--Minimal assignments were given to all students. Brighter students were encouraged

to do extra assignments. The assignments were more varied than those normally found in a homogeneous class and there was a tendency to non-verbal kinds of assignments.

2. Reading materials--While the recommendation was made that one text book be used, two of the teachers reported that they used all three.

3. Discussion--Two teachers made greater use of discussion, while one made greater use of audio-visual materials.

4. Projects--There was a greater variety of projects which were conducted by all students.

5. Group work--Students were encouraged to work in groups and to help each other, though it appears that this part of the program was not too successful.

6. Use of teacher aide--One teacher used teacher aides to work with slow students and to assist bright students in special projects.

These comments indicate that each teacher conducted his class the way it was believed would be most successful or operational. The Track One students in the heterogeneous classes scored as well as the Track One students in homogeneous classes. Track Two students in the heterogeneous classes outperformed their counterparts in the Track Two homogeneous classes. In addition, theTrack Three students trained in the heterogeneous classes obtained higher scores on the achievement tests than did the Track Three students trained in the homogeneous classes.

Appendix A. Attitude Questionnaire Employed During the First Year's Study

BERKELEY UNIFIED SCHOOL DISTRICT

INSTRUCTIONS TO STUDENTS

This is not a test. We are going to ask you many questions. These questions do not have right or wrong answers. We just want to know what the kids at Garfield think about this school and some of its classes. Since there are no right or wrong answers, answer the questions so that they are truthful for you. Remember, this is not a test.

EXAMPLE

Put an X in the space you feel is right according to you

How I feel about watching television

1.	interesting	___	___	___	___	___	___	uninteresting
2.	fun	___	___	___	___	___	___	boring
3.	educational	___	___	___	___	___	___	uneducational
4.	fast	___	___	___	___	___	___	slow

Put an X in the space you feel is right according to you

How I feel about myself

1.	lazy	___	___	___	___	___	___	hardworker
2.	honest	___	___	___	___	___	___	dishonest
3.	superior	___	___	___	___	___	___	inferior
4.	deprived	___	___	___	___	___	___	enriched
5.	black	___	___	___	___	___	___	light
6.	free	___	___	___	___	___	___	bound
7.	large	___	___	___	___	___	___	small
8.	weak	___	___	___	___	___	___	strong
9.	thick	___	___	___	___	___	___	thin
10.	deep	___	___	___	___	___	___	shallow
11.	brave	___	___	___	___	___	___	cowardly
12.	animal	___	___	___	___	___	___	human
13.	rough	___	___	___	___	___	___	smooth
14.	bully	___	___	___	___	___	___	gentle
15.	heavy	___	___	___	___	___	___	light
16.	good	___	___	___	___	___	___	bad
17.	sweet	___	___	___	___	___	___	sour
18.	equal	___	___	___	___	___	___	unequal
19.	clean	___	___	___	___	___	___	dirty
20.	rugged	___	___	___	___	___	___	delicate
21.	tasteful	___	___	___	___	___	___	distasteful
22.	hard	___	___	___	___	___	___	soft
23.	valuable	___	___	___	___	___	___	worthless
24.	beautiful	___	___	___	___	___	___	ugly
25.	pleasant	___	___	___	___	___	___	unpleasant

Put an X in the space you feel is right according to you

How I feel about myself

26.	loud	soft
27.	bitter	sweet
28.	happy	sad
29.	talkative	not talkative
30.	ferocious	peaceful
31.	failure	success
32.	light	dark
33.	nice	awful
34.	fragrant	foul
35.	healthy	sick
36.	fair	unfair
37.	active	passive
38.	kinky	straight
39.	rich	poor
40.	manly	womanly
41.	hopeful	hopeless
42.	smart	dumb
43.	slave	free
44.	fast	slow
45.	bright	dark
46.	bottom	top
47.	bored	interested
48.	hot	cold
49.	girlish	boyish
50.	sitting	moving

Put an X in the space you feel is right according to you

How I feel about the other kids at Garfield

1.	lazy	hardworker
2.	honest	dishonest
3.	superior	inferior
4.	deprived	enriched
5.	black	light
6.	free	bound
7.	large	small
8.	weak	strong
9.	thick	thin
10.	deep	shallow
11.	brave	cowardly
12.	animal	human
13.	rough	smooth
14.	bully	gentle
15.	heavy	light
16.	good	bad
17.	sweet	sour
18.	equal	unequal
19.	clean	dirty
20.	rugged	delicate
21.	tasteful	distasteful

22.	hard	___	___	___	___	___	___ soft
23.	valuable	___	___	___	___	___	___ worthless
24.	beautiful	___	___	___	___	___	___ ugly
25.	pleasant	___	___	___	___	___	___ unpleasant
26.	loud	___	___	___	___	___	___ soft
27.	bitter	___	___	___	___	___	___ sweet
28.	happy	___	___	___	___	___	___ sad
29.	talkative	___	___	___	___	___	___ not talkative
30.	ferocious	___	___	___	___	___	___ peaceful
31.	failure	___	___	___	___	___	___ success
32.	light	___	___	___	___	___	___ dark
33.	nice	___	___	___	___	___	___ awful
34.	fragrant	___	___	___	___	___	___ foul
35.	healthy	___	___	___	___	___	___ sick
36.	fair	___	___	___	___	___	___ unfair
37.	active	___	___	___	___	___	___ passive
38.	kinky	___	___	___	___	___	___ straight
39.	rich	___	___	___	___	___	___ poor
40.	manly	___	___	___	___	___	___ womanly
41.	hopeful	___	___	___	___	___	___ hopeless
42.	smart	___	___	___	___	___	___ dumb
43.	slave	___	___	___	___	___	___ free
44.	fast	___	___	___	___	___	___ slow
45.	bright	___	___	___	___	___	___ dark
46.	bottom	___	___	___	___	___	___ top
47.	bored	___	___	___	___	___	___ interested
48.	hot	___	___	___	___	___	___ cold
49.	girlish	___	___	___	___	___	___ boyish
50.	sitting	___	___	___	___	___	___ moving

We would like to know what you thought of your World History and Geography class last year. (Circle your answer.)

Remember that there are no right or wrong answers. Answer so that they are truthful for you.

1. How well did you like the classwork in World History and Geography last year?
   VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL

2. Last year the kids in World History and Geography were as smart as I am.
   ALL OF THEM MOST OF THEM FEW OF THEM NONE OF THEM

3. How much homework were you given in World History and Geography last year?
   TOO MUCH ENOUGH NOT ENOUGH NONE AT ALL

4. Last year in World History and Geography I liked to come to class.
   ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME

5. How often did you do the homework in World History and Geography last year?
   HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

6. Last year I felt impatient with some of the slow kids in my World History and Geography class.
HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

7. How much did you learn about World History and Geography last year?
QUITE A LOT A LOT A LITTLE NOTHING

8. How well did you like your teacher in World History and Geography last year?
VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL

9. Last year in World History and Geography I enjoyed doing extra work for my teacher.
HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

10. How hard did you find the work in World History and Geography last year?
VERY EASY FAIRLY EASY FAIRLY HARD VERY HARD

11. Last year I felt impatient with some of the fast kids in my World History and Geography class.
HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

12. Last year the kids in World History and Geography were like me.
ALL OF THEM MOST OF THEM A FEW OF THEM NONE OF THEM

13. Last year in World History I was happy to do the class assignments.
MOST OF THE TIME MUCH OF THE TIME ONCE IN A WHILE HARDLY EVER

14. Last year in World History and Geography discussion leaders and group members were chosen by the kids themselves.
NONE OF THE TIME SOME OF THE TIME MOST OF THE TIME ALL OF THE TIME

15. Last year in my World History and Geography class I was unable to be friendly and direct with the other kids.
NONE OF THE TIME SOME OF THE TIME MOST OF THE TIME
ALL OF THE TIME

16. How well did you like the way group leaders and members were chosen last year?
NOT VERY WELL FAIRLY WELL NOT WELL NOT AT ALL

17. Last year in World History-Geography I brought to class articles from newspapers and magazines about World History-Geography.
VERY OFTEN FAIRLY OFTEN NOT VERY OFTEN NOT AT ALL

18. Last year in World History and Geography I enjoyed working with other kids who were
A LOT SMARTER THAN I SMARTER THAN I AS SMART AS I
LESS SMART THAN I A LOT LESS SMART THAN I

19. Last year I felt at ease with the other kids in my World History-Geography class.
ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME

20. Last year I liked using the same books and materials as all of the other kids in my World History and Geography class.
VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL

21. Last year there were times when my behaviour towards the other kids in World History-Geography was quite different from the way I felt inside.
ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME

22. Last year in World History and Geography my teacher covered the book
MUCH TOO SLOWLY SLOWLY JUST RIGHT QUICKLY MUCH TOO QUICKLY

23. Last year among the other students in my World History-Geography class I respected as individuals
ALL OF THEM MOST OF THEM FEW OF THEM NONE OF THEM

We would like to know what you think of your U. S. History class this year.

Remember that there are no right or wrong answers. Answer so that they are truthful for you. (Circle your answer)

24. How well do you like the classwork in U. S. History this year?
VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL

25. This year the kids in U. S. History were as smart as I am
ALL OF THEM MOST OF THEM FEW OF THEM NONE OF THEM

26. How much homework is given in U. S. History this year?
TOO MUCH ENOUGH NOT ENOUGH NONE AT ALL

27. This year in U. S. History I like to come to class
ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME

28. How often do you do the homework in U. S. History this year?
HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

29. This year I feel impatient with some of the slow kids in my U. S. History class
HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

30. How much are you learning about U. S. History this year?
QUITE A LOT A LOT A LITTLE NOTHING

31. How well do you like your teacher in U. S. History this year?
VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL

32. This year in U. S. History I enjoy doing the extra work for my teacher.
HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

33. How hard do you find the work in U. S. History this year?
VERY EASY FAIRLY EASY FAIRLY HARD VERY HARD

34. This year I feel impatient with some of the fast kids in my U. S. History class.
HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

35. This year the kids in U. S. History are like me.
ALL OF THEM MOST OF THEM A FEW OF THEM NONE OF THEM

36. This year in U. S. History I am happy to do the class assignments.
MOST OF THE TIME MUCH OF THE TIME ONCE IN A WHILE HARDLY EVER

37. This year in U. S. History discussion leaders and group members are chosen by the kids themselves.
NONE OF THE TIME SOME OF THE TIME MOST OF THE TIME ALL OF THE TIME

38. This year in my U. S. History class I am unable to be friendly and direct with the other kids.
NONE OF THE TIME SOME OF THE TIME MOST OF THE TIME ALL OF THE TIME

39. How well do you like the way group leaders and members are chosen this year?
NOT VERY WELL FAIRLY WELL NOT WELL NOT AT ALL

40. This year in U. S. History I bring to class articles from newspapers and magazines about U. S. History.
VERY OFTEN FAIRLY OFTEN NOT VERY OFTEN NOT AT ALL

41. This year in U. S. History I am enjoying working with other kids who are
A LOT SMARTER THAN I SMARTER THAN AS SMART AS I LESS SMART THAN I A LOT LESS SMART THAN I

42. This year I feel at ease with the other kids in my U. S. History class.
ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME

43. This year I wanted to use the same books and materials as all of the other kids in my U. S. History class.
NOT VERY MUCH NOT MUCH MUCH VERY MUCH

44. This year there are times when my behavior towards the other kids in U. S. History is quite different from the way I feel inside.
ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME

45. This year in U. S. History my teacher is covering the book
MUCH TOO SLOWLY SLOWLY JUST RIGHT QUICKLY MUCH TOO QUICKLY

46. This year among the other students in my U. S. History class I respect as individuals
ALL OF THEM MOST OF THEM FEW OF THEM NONE OF THEM

47. In a group that I choose myself, I learn MORE THAN SAME AS LESS THAN I would learn in a group chosen by a teacher

Appendix B: Letters sent to parents asking them to volunteer their children for participation in the study.

GARFIELD JUNIOR HIGH SCHOOL
BERKELEY, CALIFORNIA

Dear Parents:

As you know, your child has been in a social science class-heterogeneous grouping study for the past year.

We are pleased to report that the study has now been funded by the United States Office of Education (USOE) so that we will be able to continue the study for one more year at West Campus.

In addition, Mr. Eugene McCreary, supervisor to teacher education (social sciences), of the University of California has agreed to be consultant to the teachers of the heterogeneous groups for the duration of the study.

We prefer not to introduce new students into the study but would like to continue with the students already in the study. If you would like your child to have the opportunity of continuing please let me know before June 7, 1968.

If you have any questions, Mr. Wright, the Project Director, is available to answer them. You can call him at 841-1422, extension 353.

Sincerely yours,

/s/ F. Carl Dwight
Principal, West Campue

____Please include my child in the grouping experiment at West Campus.

____Please do not include my child in the grouping experiment at West Campus.

Parent's Signature _____

Student's Name _____

GARFIELD AND WILLARD JUNIOR HIGH SCHOOLS
BERKELEY, CALIFORNIA

Dear Parents:

We have received a grant from the United States Office of Education (USOE) to conduct a study in heterogeneous grouping in World History at West Campus. As a result we would like to set up heterogeneous and homogeneous classes at West Campus.

We would like your child to participate in this valuable study. If you would like your child to have the opportunity of participating in this experiment please let me know by June 7, 1968. Unfortunately, even if you consent to have your child participate we have a limited number of places, so your child may not be assigned to a heterogeneous group.

As a supplementary aid, Mr. Eugene McCreary of the University of California, supervisor of teacher education (social sciences), has agreed to be consultant to the teachers of the heterogeneous groups for the duration of the study.

If you have any questions, Mr. Wright, the Project Director, is available to answer them. You can call him at 841-1422, extension 353.

Sincerely yours,

/s/ F. Carl Dwight
Principal, West Campus

____Please include my child in the grouping project at West Campus.

____Please do not include my child in the grouping experiment at West Campus.

Parent's signature_____

Student's Name_____

Appendix C. Second Semester Pre Test.

WEST CAMPUS BERKELEY HIGH SCHOOL, HISTORY DEPARTMENT

"MODERN TIMES"
Second Semester Pre Test

Multiple Choice. Select the best answer for each question.

1. One who favors changes to remedy the evils of society, but does not favor radical or fast changes is called a:
a) liberal, b) conservative, c) rationalist, d) proletariat

2. A writ of habeas corpus guarantees each person
a) freedom of the press, b) protection against excessive taxes, c) the right to know why he is being arrested, d) the right to serve on a jury.

3. In almost all revolutions, which is generally not an aim at the outset (at first):
a) overthrow the government, b) to end injustice and inequality, c) to end waste and corruption, d) to gain a voice in the government.

4. Which revolution occurred first:
a) French, b)American, c) English, d) Russian.

5. Many kings came into conflict with their people because the kings claimed the right to rule by:
a) popular consent, b) divine right, c) the will of the elected representatives, d) the will of the Church.

6. Parliament, Kuomintang, Estates-General, and Duma all refer to:
a) political parties, b) law making groups, c) villages, d) ruling leaders.

7. The term "democracy" means government by:
a) the will of the states, b) most of the people, c) best educated, d) most wealthy.

8. Which major foreign nation aided the American colonies in their revolution against England.
a) Austria, b) Russia, c)Germany, d)France.

9. The Bill of Rights signed by William and Mary:
a) increased the power of the king over Parliament, b) increased the power of the Parliament over the king, c) recognized the "divine right" of kings, d) established the right of the king to veto laws.

10. Which was not a cause of the French Revolution:
a)the power and privileges of the Church, b) social unrest and misery, c) tyrannical and absolute kings, d) liberal kings.

11. The French method of execution during the Revolution was by:
a) guillotine, b) hanging, c) electrocution, d) gas.

12. What was the slogan of the French Revolution:
a) Liberty, equality and the pursuit of happiness, b) Liberty equality, fraternity, c) Brotherhood of man, d) Peace, bread, and land.

13. Who was the father of Bolshevism?
a) Lenin, b)Stalin, c) Karensky, d) Peter the Great.

14. Which was not a reason which allowed the Communists to take over in Russia in 1917.
a) need for bread and more to eat, b) need for land distribution, c) desire for peace, d) military successes in World War I.

15. Which is true of the Communist party in Russia?
a) most citizens are members, b) the people who control the Party control Russia, c) Party members are free to obey or disobey Party policies, d) citizens may enter or leave the Party at will.

16. The Nationalist Chinese government on Taiwan is today headed by:
a) Sun Yat Sen, b) Mao Tse Tung, c) Chiang-Kai-Shek, d) Chou en lai.

17. The "great leap forward" refers to Communist China's effort to increase
a) its population, b) amount of land under its control, c) number of men in its armies, d) output of agriculture and other goods.

18. The last nation to unite was:
a) England, b) Russia, c) France, d) Germany

19. Which of the following peoples made the least important contribution to the people of Spain?
a) Goths, b) Romans, c) Moors, d) Mongols.

20. Which of the following statements is least characteristic of European absolutism during the 17th and 18th centuries?
a) the Church exercised great restraint over royal power, b) kings were thought to rule as divine agents, c) strong

monarchies were thought to be necessary to maintain civil peace, d) absolutist kings often led their nations into war for prestige and power.

21. French absolutism reached its greatest strength and splendour under:
a) Henry IV, b) Louis XIV, c) Louis XIII, d) Louis XVI.

22. With which scientific theory is Charles Darwin associated?
a) theory of evolution, b) theory of the atom, c) theory of heredity, d) theory of relativity.

23. The first series of inventions in the Industrial Revolution occurred in:
a) ship building, b) mining, c) agriculture, d)production of cloth.

24. In the Wealth of Nations, Adam Smith pointed out the advantages of the:
a) doctrine of the divine right of kings, b) principle of separation of power, c) principle of freedom of trade, d) theory of mercantilism.

25. Which of the following is not an advantage of specialization of labor in industry?
a) production of goods is much faster, b) greater efficiency may result in lower prices, c) less highly skilled working force may be employed, d) the employee takes greater interest and pride in the finished article.

26. The Industrial Revolution came first to:
a) Britain, b) France, c) Germany, d)United States.

27. Socialism means:
a) abolition of factory managers, b) abolition of all government, c) ownership of land by the peasants, d) ownership of industry by the government or community.

28. Which does not correctly sum up the theory of Marx?
a) workers will revolt, b) workers will rise to middle class status, c) private industry will be abolished, d) workers will triumph.

29. Nationalism is best defined as a feeling of:
a) loyalty to the king, b) loyalty to and pride in a country, c) loyalty to a legislative body, d) comradeship with the leaders.

30. Royal power was increased in medieval France when:
a) strong local leaders established their authority after Charlemagne's death, b) King established strong French parliament, c) Joan of Arc helped same French throne, d) Louis VI lost power.

31. The Hundred Years War between England and France was caused by:
a) rivalty over territories and trade, b) religious differences, c) papal interference, d) disputes over colonies.

32. Napolean gained victory after victory until he attacked:
a) Italy, b) Spain, c) Russia, d) Austria.

33. In 1815 the powers of Europe gathered at Vienna to:
a) prepare trial of Napolean, b) restore the old bounderies and old royal families, c) arrange for democratic principles in Europe, d) set up a League of Nations.

34. Victor Hugo wrote the world-famous novel:
a) War and Peace, b) Oliver Twist, c) Les Miserables, d) The Thinker.

35. Pierre and Marie Curie discovered:
a) radium, b) uranium, c) iron, d) zinc.

36. Louis XIV of France and James I of England both believed in:
a) teachings of Martin Luther, b) divine right of kings, c) advantages of democracy, d) right of commoners to free public education.

37. In the fourteenth century the English Parliament gradually won control of the government by:
a) waging war against the king, b) demanding privileges in return for granting new taxes, c) terms of Magna Carta, d), increasing membership of the House of Lords.

38. The Industrial Revolution began in England because of:
a) abundance of raw materials, money and skills, b) favorable climate, c) the help of the royal family, d) large number of good roads.

39. The unit of measure for an electric current is named for its inventor:
a)Marconi, b) Torricelli, c) Volta, d) Fermi.

40. The reformation left Germany:
a) divided religiously, b) all Catholic, c) all Protestant, d) with religious freedom.

41. The family that ruled Prussia was:
a) Hapsburgs, b) Hohenzollerns, c) Capets, d) Tudors.

42. After World War II Germany:
a) was divided into two countries, b) became all Communist, c) was absorbed by her neighbors, d) was given a U. N. mandate in Africa.

43. The great contribution of Henry Ford to manufacturing was:
a) the assembly-line, b) gasoline motor, c) self-starter, d) automation.

44. The United Stated entered World War I in 1917 because: a) a German army was transferred to Mexico, b) German subs put American lives and property in danger, c) the Allies promised to support a League of Nations, d) German mis-treatment of Jewish citizens.

45. Immediately after World War I, the United States: a) took an active role in European affairs, b) joined the League of Nations, c) concentrated on its own affairs, d) helped establish a republic in Germany.

46. Peter the Great was determined to: a) conquer Poland, b) make Moscow a cultural center, c) westernize Russia, d) convert the Russians to Catholicism.

47. Gandhi urged a policy of: a) guerrilla warfare, b) cooperation with the British, c) passive resistance, d) alliance of all Third World forces against Europe.

48. British take-over of India is an example of: a) Communism, b)Imperialism, c) Socialism, d) Nationalism.

49. The richest prize in the Empire of the Dutch was: a)Cambodia, b) Laos and Viet Nam, c) the East Indies, d) Philippines.

50. A "sphere of influence" was: a) a port where all Europeans could trade equally, b) an area where Christian missioners were given special treatment, c) an area where only one power had special rights, d) a port which was controlled by the British.

51. In the 19th Century, China: a) was a powerful military nation, b) enjoyed a Golden Age, c) was divided into "spheres of influence", d) was conquered by Japan.

52. "I disagree with what you say, but I will fight to the death for your right to say it." Which of the following men might have said this: a) John Calvin, b) Oliver Cromwell, c) Voltaire, d) Louis XIV.

53. The earliest English laws to improve labor conditions: a) showed concern for women and children, b) dealt with labor unions, c) passed by the Labor Party, d) provided for government control of industry.

54. One affect of the agricultural revolution was: a) to change farming from a way of life to a business, b) increase the number of farm workers, c) encourage small family-size farms, d) decrease the farmers need for bank loans.

55. Bookkeeping was important to the new middle class because it:
a) increased production, b) added profits to business, c) kept track of costs and profits, d) did away with taxes.

56. Workers joined early unions:
a) to own property, b) to gain the right to vote, c) to get higher pay, d) to gain political power.

57. Which event came first:
a) World War I, b) Communish Revolution in Russia, c) French Revolution, d) Formation of the United Nations.

58. Who were contemporaries (lived at the same time):
a) Hitler, Stalin, b) Napolean, Lenin, c) da Vinci, Picasso, d) F. Roosevelt, Louis XIV.

59. Baroque architecture is characterized by its:
a) lavish ornamentation, b) flying buttresses, c) functional design, d) skyscraper appearance.

60. For the common people the most important advantage of feudalism (Middle Ages) was:
a) material wealth, b) intellectual stimulation, c) opportunities for adventure, d) secure place in society.

Appendix D. Spring Post Test

BERKELEY UNIFIED SCHOOL DISTRICT

This is not a test. The questions do not have right or wrong answers. We just want to know what the students think about this school and some of its classes. Since there are no right or wrong answers, answer the questions so that they are truthful for you. Do not skip any questions. Check the answer which is closest to what you feel is true. Remember this is not a test.

Example. Check the space on the blanks between the pairs of words which is closest to your feelings. Make your answer mark clearly. Make sure you check only one space for each pair of words.

How I feel about television.

1. interesting ___ ___ ___ ___ ___ ___ uninteresting

A check in the first space represents your feeling that television is most interesting or very, very interesting. A check in the last space represents your feeling that television is very, very uninteresting.

2. heavy ___ ___ ___ ___ ___ ___ light

This pair of adjectives might refer to the quality of the programs you think about or it might refer to how watching TV makes you feel, whether it depresses you or makes you feel good. You must decide how the words apply to the subject and then decide how you feel about that way of looking at it.

Put an X in the space you feel is right for you

How I feel about myself

1.	honest	___	___	___	___	___	___	dishonest
3.	weak	___	___	___	___	___	___	strong
5.	cowardly	___	___	___	___	___	___	brave
7.	fast	___	___	___	___	___	___	slow
9.	bitter	___	___	___	___	___	___	sweet
11.	bully	___	___	___	___	___	___	gentle
13.	distasteful	___	___	___	___	___	___	tasteful
15.	sick	___	___	___	___	___	___	healthy
17.	fair	___	___	___	___	___	___	unfair
19.	ugly	___	___	___	___	___	___	beautiful
21.	clean	___	___	___	___	___	___	dirty
23.	passive	___	___	___	___	___	___	active
25.	happy	___	___	___	___	___	___	sad
27.	dumb	___	___	___	___	___	___	smart
29.	valuable	___	___	___	___	___	___	worthless
31.	bottom	___	___	___	___	___	___	top
33.	good	___	___	___	___	___	___	bad
35.	unpleasant	___	___	___	___	___	___	pleasant
37.	nice	___	___	___	___	___	___	awful
39.	hopeful	___	___	___	___	___	___	hopeless

How I feel about the other kids in this school

2.	honest	___	___	___	___	___	___	dishonest
4.	weak	___	___	___	___	___	___	strong
6.	cowardly	___	___	___	___	___	___	brave
8.	fast	___	___	___	___	___	___	slow
10.	bitter	___	___	___	___	___	___	sweet
12.	bully	___	___	___	___	___	___	gentle
14.	distasteful	___	___	___	___	___	___	tasteful
16.	sick	___	___	___	___	___	___	healthy
18.	fair	___	___	___	___	___	___	unfair
20.	ugly	___	___	___	___	___	___	beautiful
22.	clean	___	___	___	___	___	___	dirty
24.	passive	___	___	___	___	___	___	active
26.	happy	___	___	___	___	___	___	sad
28.	dumb	___	___	___	___	___	___	smart
30.	valuable	___	___	___	___	___	___	worthless
32.	bottom	___	___	___	___	___	___	top
34.	good	___	___	___	___	___	___	bad
36.	unpleasant	___	___	___	___	___	___	pleasant
38.	nice	___	___	___	___	___	___	awful
40.	hopeful	___	___	___	___	___	___	hopeless

We would like to know what you thought of your social studies class this year. (Circle your answer)

Remember that there are no right or wrong answers. Answer so that they are truthful for your.

1. How well did you like the classwork in social studies last year?
   VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL

2. Last year the kids in social studies were just as smart as I am.
   ALL OF THEM MOST OF THEM FEW OF THEM NONE OF THEM

3. Last year in social studies I liked to come to class.
   ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME

4. Last year I felt impatient with some of the slow kids in social studies class.
   HARDLY EVER ONCE IN A SHILE MUCH OF THE TIME MOST OF THE TIME

5. How much did you learn about social studies last year?
   QUITE A LOT A LOT A LITTLE NOTHING

6. How well did you like your teacher in social studies last year?
   VERY WELL FAIRLY WELL NOT VERY WELL NOT AT ALL

7. Last year in social studies I enjoyed doing extra work for my teacher.
   HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

8. Last year the smart kids getting most of the attention made me mad.
   HARDLY EVER ONCE IN A WHILE MUCH OF THE TIME MOST OF THE TIME

9. Last year in social studies I was happy to do the class assignments.
   MOST OF THE TIME MUCH OF THE TIME ONCE IN A WHILE HARDLY EVER

10. Last year in social studies I brought to class articles from newspapers and magazines.
    VERY OFTEN FAIRLY OFTEN NOT VERY OFTEN NOT AT ALL

11. Last year I could be myself with the other kids in my social studies class.
    ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME

12. Last year in social studies my teacher covered the book
MUCH TOO SLOWLY SLOWLY JUST RIGHT QUICKLY MUCH TOO QUICKLY

13. Last year in social studies I acted differently from the way I felt inside.
ALL OF THE TIME MOST OF THE TIME SOME OF THE TIME NONE OF THE TIME

14. Last year in social studies I found I could respect
ALL OF THE STUDENTS MOST OF THE STUDENTS FEW OF THE STU_ DENTS NONE OF THE STUDENTS

Now we would like to know what the adults at home thing about your school.

1. The adults at home think I'm learning things in school which will help me later on in life.
TRUE FALSE

2. The adults at home think my counselors are helpful in planning my school program.
TRUE FALSE

3. The adults at home think there should be special classes for smart kids.
TRUE FALSE

4. The adults at home think this school is below standard.
TRUE FALSE

5. The adults at home think there should be no tracking in social studies.
TRUE FALSE

6. The adults at home think my teachers treat me fairly.
TRUE FALSE

7. The adults at home think that all the kids should use the same books and materials.
TRUE FALSE

8. The adults at home think that my counselors are helpful in handling school problems.
TRUE FALSE

9. The adults at home think there should be tracking in Math and English.
TRUE FALSE

10. The adults at home think my teachers are well prepared to teach their classes.
TRUE FALSE

11. The adults at home think I should go to a different school.
TRUE FALSE

12. The adults at home think there should be speciall classes for kids who are slow at learning.
TRUE FALSE

13. The adults at home think my school will help me learn how to get along with other people.
TRUE FALSE

14. The adults at home think all the kids in a class should be of the same ability.
TRUE FALSE

15. The adults at home think my school is a waste of time.
TRUE FALSE

REFERENCES

1. Wyndham, H. S., Ability Grouping. Melbourne University Press, Oxford University Press, 1934.

2. Ekstrom, R., Experimental Studies of Homogeneous Grouping: A Review of Literature. Princeton: Educational Testing Service, 1959.

3. Conant, J. B., The American High School Today. New York: McGraw-Hill Book Co., 1959, pp. 49-50.

4. Otto, H. J., "Homogeneous Grouping," in W. Monroe (ed.), Encyclopedia of Educational Research. New York: Macmillan Co., 1950, pp. 376-378.

5. Rock, R. T., Jr., "A Critical Study of Current Practices in Ability Grouping," Educational Research Bulletin, Catholic University of America, IV, Nos. 4 and 5, 1929.

6. Rankin, P., "Pupil Classification and Grouping," Review of Educational Research, I, 1931, pp. 200-230.

7. Turney, A. H., "Status of Ability Grouping," Educational Administration and Supervision, XVII, 1931, pp. 21-42, 110-127.

8. Billett, R. O., Provisions for Individual Differences, Marking and Promotion, Bulletin 1932, No. 17, Monograph 13. Washington: U. S. Office of Education, 1932.

9. Cornell, Ethel L., "Effects of Ability Grouping Determinable from Published Studies," in The Grouping of Pupils, Thirty-fifth Yearbook of the National Society for the Study of Education, Part I. Chicago: Distributed by the University of Chicago Press, 1936, pp. 289-302.

10. Luttrell, Jack S., Doctoral Dissertation, University of North Carolina, 1958, 139 pages.

11. Schaefer, Vincent J., "The Boy Who Learned to Grow Salt," Saturday Review, December 12, 1959, p. 54.

12. Koontz, William F., A Study of Achievement as a Function of Homogeneous Grouping. J. Exp. Ed. XXX, 1961, pp. 249-253.

13. Drews, Elizabeth M., A Report to the Lansing, Michigan, Public Schools, 1962. Unpaged, Dittoed.

14. Atkinson, John W., and O'Connor, Patricia. Effects of Ability Grouping in Schools Related to Individual Differences in Achievement-Related Motivation. Cooperative Research Project No. 1283. Washington: U. S. Department of Health, Education, and Welfare, Office of Education, March 1963, 164 pages.

15. Borg, Walter R., An Evaluation of Ability Grouping. Cooperative Research Project No. 577, Utah State University, 1964, 441 pages.

16. Goldberg, Mirian L., Passow, A. Harry, and Justman, Joseph, The Effects of Ability Grouping. Horace Mann-Lincoln Institute of School Experimentation Talented Youth Project, Teachers College, Columbia. Teachers College Press, New York.

17. Dockrell, W. B., Edmonton Junior High School Streaming Project. Studies in Grouping, Alberta Teachers Association, in press.

18. Coleman, James S., Equality of Educational Opportunity. Johns Hopkins University, U. S. Department of Health, Education and Welfare, p. 205.

19. Scheffé, H. Analysis of Variance. New York: John Wiley & Sons.